Arthaniti

(New Series)

EDITORIAL BOARD

Chairman
Soumyen Sikdar

Joint Editors
Sarbajit Chaudhuri
Sankar Kumar Bhaumik

Associate Editors
Ishita Mukhopadhyay
Rabindranath Mukhopadhyay

Book Review Editors
Nita Mitra
Anjan Chakraborty

Editorial Secretary
Kanai Lal Basak

Members
Asis Kumar Banerjee
Arup Mallick
Ajit Chaudhuri
Kalyan Sanyal
Sarmila Banerjee
Pabitra Kumar Giri
Manash Ranjan Gupta
Sugata Marjit
Dipankar Coondoo
Abhirup Sarkar

Editorial Advisory Board
Ronald W Jones
Timothy Besley
Sajal Lahiri
Partha Sen
Prasanta K. Pattanaik
Rabindra Nath Bhattacharya
Prabuddha Nath Roy
Amiya K Bagchi
Dilip Mookherjee
Pasquale Sgro
Debraj Ray
Hiron Kumar Banerjee
David E A Giles
Bharat R Hazari
Hamid Beladi
Graciela Chichilnisky
Ira Gang
Shigemi Yabuuchi
Ulaganathan Sankar

Volume II, Nos. (1-2),2003

Contents

Deregulation, Ownership and Efficiency in Indian Banking
Subal C. Kumbhakar and Subrata Sarkar

Court Delays and the Defendant Behaviour : An Economic Analysis
Andrew Torre and Bharat R. Hazari

Frontier Production Function and Extreme Observations
M. Pal, B. Ghosh and C. Neogi

On Rationalization of Nash and Subgame-Perfect Behaviour from Observed Outcomes
Indrajit Ray and Susan Snyder

Foreign market entry and host-country welfare : A theoretical analysis
Arijit Mukherjee

Child Labour and Economic Development : A Theoretical Analysis
Manash Ranjan Gupta

When scrutiny reduces signaling incentives
Debabrata Datta and Jaideep Roy

Book Review

EDITORIAL/ BUSINESS OFFICE
Department of Economics
University of Calcutta
56A, B.T. Road, Kolkata 700 050, India.
Tel: 91-33-2557-5082, 91-33-2528-4066
Fax: 91-33-2528-4066
E-mail: sarbajitch@yahoo.com
spsikdar@cal3.vsnl.net.in

ARTHANITI

(NEW SERIES)

THE JOURNAL OF THE DEPARTMENT OF ECONOMICS, UNIVERSITY OF CALCUTTA

Aims & scope: The Dept. of Economics, University of Calcutta is resuming the publication of its journal *ARTHANITI* in 2002 as a new series. For the time being, *ARTHANITI* will be published as an annual international journal. Papers across the entire field of economics, including econometrics, economic history, and political economy will be considered for publication. Both theoretical and empirical manuscripts are equally welcomed.

Manuscripts will be judged solely on the quality of their content. *ARTHANITI* will adopt a double-blind review process. The editorial board will not reveal the identities of either authors or referees to other parties in the review process. The editor and/or any member of the editorial board will send submitted papers to the referees only after an initial reading. The editorial board reserves the right to reject any paper at that stage. However, most of the submitted papers will be sent to the referees for comments.

G165632

SUBSCRIPTION RATES:

Indian. Institutions:	Rs. 300/- (Annual)	Overseas: Institutions:	US $ 40 (Annual)
	Rs. 750/- (3 Years)		US $ 100 (3 Years)
Individuals:	Rs. 150/- (Annual)	Individuals:	US $ 25 (Annual)
	Rs. 375/- (3 Years)		US $ 60 (3 Years)
		Air Surcharge:	US $ 5 (optional)

All rights reserved. No part of the contents of this journal may be reproduced in any form without the written consent of the publisher.

Published by Soumyen Sikdar, Chairman, Editorial Board of ARTHANITI, on behalf of the Department of Economics, Calcutta University.

Printed at· Calcutta University Press

Computer Type Setting & Designing : Suchhanda Chaudhury

Front Page Design : Suchhanda Chaudhury

cuk-H00558-2-u71b5b82.

ARTHANITI

New Series

Volume II, Nos. (1-2), 2003

133/2

Contents

Deregulation, Ownership and Efficiency in Indian Banking
Subal C. Kumbhakar and Subrata Sarkar 1-26

Court Delays and the Defendant Behaviour : An Economic Analysis
Andrew Torre and Bharat R. Hazari 27-41

Frontier Production Function and Extreme Observations
M. Pal, B. Ghosh and C. Neogi 42-61

On Rationalization of Nash and Subgame-Perfect Behaviour from Observed Outcomes
Indrajit Ray and Susan Snyder 62-78

Foreign market entry and host-country welfare : A theoretical analysis
Arijit Mukherjee 79-96

Child Labour and Economic Development : A Theoretical Analysis
Manash Ranjan Gupta 97-107

When scrutiny reduces signaling incentives
Debabrata Datta and Jaideep Roy 108-118

Book Review 119-121

Arthaniti 2 (1-2), 1-26 / 2003

Deregulation, Ownership and Efficiency in Indian Banking*

Subal C. Kumbhakar** and Subrata Sarkar***

Abstract

This paper uses a Stochastic Cost Frontier Approach to evaluate the efficiency of the Indian banking system using panel data on public and private sector banks for the period 1986-2000. Efficiency variations in terms of exogenous factors are used to analyze the time behavior of the banking system, especially the changes in efficiency since the initiation of the reforms program in 1992 Ownership characteristics of banks are also incorporated into the analysis to examine if efficiency as well as efficiency changes have differed across ownership groups.

Empirical results indicate the presence of cost inefficiency in the Indian banking system, but there is a tendency for inefficiencies to decline over time. The results also indicate that cost inefficiency of banks has increased since the initiation of the reforms, though the reduction in inefficiencies over time continues albeit at a slower rate compared to that observed in the pre-deregulation period. We also find that the private banks are generally more cost-efficient than public banks, but there are no significant differences in the impact of deregulation on the cost efficiency of these two bank groups. At the individual level, we find marked differences in the efficiency behaviour of different banks with private banks exhibiting much more intra-group volatility in relative efficiency changes between the pre and post deregulation periods compared to that of public banks.

JEL Classification : C23, G21.

Keywords : Stochastic Cost Frontier, Deregulation, Banking Efficiency.

1. Introduction

In 1991, based on the recommendations of the Narasimham Committee on Financial Sector Reforms, the Reserve Bank of India (RBI) launched major banking sector reforms aimed towards improving the performance of the India banking system. These reforms, which were a part of the broader program of globalization of the Indian economy, encompassed entry deregulation, branch de-licensing, deregulation of interest rates, and mandating strong public sector banks to go to the capital market to raise funds up to 49 percent of their equity capital. The reforms also entailed a gradual reduction of the Cash Reserve Ratio (CRR) and the Statutory Liquidity Ratio (SLR), and a strengthening of the banking system through the institution of the Bank of International Settlements (BIS) norm of an 8 percent Capital Adequacy Ratio, as well as stringent income recognition and provisioning norms (*see* Sarkar (1999) for an exhaustive review of the recent banking sector reforms). These changes were aimed

---

* We thank Rudra Sensarma and Arijit Ghosh for excellent research assistance. None, other than us, is responsible for remaining errors.

** Department of Economics, State University of New York, Binghamton, NY 13902, USA. E mail : kkar@binghamton.edu

*** Indira Gandhi Institute of Development Research, Gen. Vaidya Marg, Goregaon (East), Mumbai 400 065, India, E mail : ssarkar@igidr.ac.in

towards creating a competitive environment that, in the long run, was expected to lead to substantial gains in efficiency, profitability, and productivity of the banking system as a whole, and that of the public banks, in particular.

In this paper we evaluate the efficiency of the Indian banking system using panel data on public and private sector banks for the period 1986-2000. The long period of the panel, encompassing both the pre-deregulation period (1986-1992) and the post-deregulation period (1993-2000) enables us to examine econometrically if the time behaviour of efficiency shows any structural break over these two periods, thereby providing us a statistical means of judging the effect of the reforms on banking efficiency. In addition, since the Indian banking system has commercial banks belonging to both public and private sectors, we also examine whether efficiency varies across ownership groups, and whether the responses to the liberalization measure is different for different ownership groups. We thus contribute to the ownership and performance literature. For both these analyses, we use stochastic cost frontier models that allows us not only to estimate producer specific efficiencies but also to explain their variations in terms of exogenous factors. Finally, since the frontier analysis gives efficiency estimates for each bank, the analysis enables us to identify those banks that are in need of relatively more policy intervention for increasing the efficiency of their operations.

Stochastic frontier analysis has been used by a number of studies in evaluating banking efficiency, for example by Berger and Mester (2001) with respect to U.S. banking; by Mendes and Reblo (1999) with respect to Portuguese banking; by Chaffai (1997) with respect to Tunisian banking; by Grifell-Tatje and Lovell (1996), and Kumbhakar et al.(2001) with respect to Spanish banking; and by Berg, et al. (1993), for Nordic countries, to name a few. With respect to India, while there are quite a few studies that have looked into the performance and productivity of the Indian banking system (Subramanyam (1993); Bhattacharyya, Lovell and Sahay (1996); Bhattacharyya, Bhattacharyya and Kumbhakar (1997); Kumbhakar and Sarkar (2003)), almost all of these studies were based on the usual average response function. In other words, these studies assumed all banks to be efficient so that the only deviation of output or cost from the maximum (minimum) attainable level, was due to purely random factors. However, previous empirical work (*see* Berger and Humphrey (1992) and references therein) has demonstrated that there are often wide variations in the performance of individual banks within the banking industry. If this is the case, then the stochastic frontier analysis framework is more suitable for analyzing banking performance, and accordingly we adopt this approach in our analysis.

The rest of the paper is organized as follows. In Section 2 we present a brief overview of the institutional structure and the regulatory environment of Indian banking and the recent deregulation measures. Section 3 contains a short discussion of stochastic frontier analysis and the models that we estimate in the paper. Section 4 describes the data used and the econometric models estimated. Section 5 presents the empirical findings and Section 6 concludes.

2. Institutional Structure and Regulatory Environment of Indian Banking

The banking system in India, like those in most developing economies, is characterized by the co-existence of diffecent ownership groups, public and private, and within private, domestic and foreign. Public sector banks in India came into existence in several phases. In 1955, the Government of India (GOI) took over the ownership of the Imperial Bank of India and reconstituted it as the State Bank of India (SBI) under the State Bank of India Act of 1955, Later, in 1959, the State Bank of India (Subsidiary Banks) Act was passed enabling the SBI to take over seven banks of princely states as its associate banks. The SBI and its associates ware entrusted with the task of serving the banking needs of the hitherto neglected sectors. However, notwithstanding the progress made by these banks in terms of geographical coverage and credit expansion, it was felt that commercial bank credit was flowing mainly to the large and well established business houses, while sectors such as agriculture and small scale industries were being neglected. Thus in 1967, the policy of social control over banks was announced, and in 1969, fourteen of the largest private banks were nationalized under the Nationalization Act of 1969. A second phase of bank nationalization followed with six more private banks getting nationalized in 1980. The smaller private banks as well as the foreign banks were allowed to coexist with the public sector banks but their activities were highly restricted through entry regulation and strict branch licensing policies.

With the nationalization of the major commercial banks, a large number of regulatory measures were adopted by the RBI. Apart from changing the sectoral composition of credit, the RBI stipulated lending targets to priority sectors, provided refinancing facilities, set up credit guarantee schemes, and directed banks to open branches in rural and semi-urban areas to make banking accessible to all. The RBI also fixed maximum deposit rates on both savings and time deposits of all maturities and specified differential lending rates linked to borrowers'income and types of lending. The Lead Bank Scheme was started for designing and implementing credit plans at the micro level.These measures led to the phenomenal growth of the banking system in general, and the public sector banks, in particular. By the early nineties, public sector banks accounted for nearly 90 percent of total deposits and advances, with the residual being almost equally split between private and foreign banks (Table 1).

However, by this time, the excessive focus on quantitative achievements had made many of the public sector banks unprofitable and under-capitalized by international standards. Many banks were earning less than reasonable rates of return, had low capital adequacy and high non-performing assets, and were providing poor quality customer service.Operating costs were increasing at a very high rate and the rapid growth in staff and promotions had diluted the quality of manpower.

In recognition of these growing illnesses, the RBI launched major banking sector reforms in 1991 aimed at creating a more profitable, efficient and sound banking system, based on the recommendations of the first Narasimham Committee on Financial Sector Reforms. The reforms sought to improve bank efficiency through entry deregulation, branch de-licensing, deregulation of interest rates, and mandating strong public sector banks to go to the capital market to raise funds upto 49 percent of their equity capital. The last move was primarily aimed towards generating market pressures on good public sector banks so that they became more efficient. The reforms also targeted to improve bank profitability through the gradual reduction of the CRR and the SLR, and to strengthen the banking system through the institution of the Bank of International Settlements (BIS) norm of an 8 percent Capital Adequacy Ratio (CAR), as well as stringent income recognition and provisioning norms.

These changes were aimed towards creating a competitive environment that, in the long-run, was expected to lead to substantial gains in efficiency, profitability and productivity. Signs of increased competition has appeared in the Indian banking industry, as is evident in the decline of the four-bank asset concentration ratio from 0.49 in 1991-92 to 0.44 in 1994-95, by the growing presence of the private and foreign banks, and in the appearance of service competition.The performance of public banks has also become more market-driven with growing emphasis put on profitability as an important benchmark for evaluating their performance by policy makers in the post-reforms era (MOF, 1993).

The existence of regulatory bottlenecks in the past, and their gradual liberalization in the recent years, provide us with a natural experiment that is well suited for studying the effects of deregulation on efficiency changes. In addition, the Indian banking system which consists of commercial banks belonging to both public and private sectors, is particularly well suited for examining whether efficiency change could vary across ownership groups.

3. Stochastic Cost Frontier Approach

The choice of an appropriate behavioural assumption is crucial in any economic modeling. The most widely accepted behavioural assumption for service industries is that of cost minimization, in which outputs are treated as exogenous. We follow the other banking studies and assume that an individual bank's objective is to minimize cost. Thus the optimization problem for a typical bank can be written as :

$$\text{Min.}\, c = w'x \text{ subject to } F(y,\theta.x,t) = 0 \qquad (1)$$

where $w=(w_1,...,w_J)'$ is a $(J \times 1)$ vector of variable input prices; $x=(x_1,...,x_J)'$ is a $(J x 1)$ vector of variable inputs; $y=(y_1,...,y_M)'$ is an $(M x 1)$ output vector; t is a trend variable capturing technical change; and $F(.)$ is the production possibility frountier. Finally, $\theta \leq 1$ is a scalar that denotes input-oriented technical efficiency (Farrell (1957)). Thus, although a bank uses the input vector x, in effective units it is worth $\theta \cdot x \leq x$. Alternatively, to produce a given vector of output, an inefficient bank uses each and every input by $(1-\theta)100\%$ more. Note that this formulation of technical inefficiency is different from the one used in Aigner et al. (1977) and Meeusen and van den Broeck (1977), which is labeled as output-oriented technical inefficiency.

The solution of the above problem gives the conditional input demand functions, viz.,

$$x_j = x_j(w,y,t)/\theta\ ,$$

where $x_j(w,y,t)$ is the conditional input demand for input j for a fully efficient bank. Using the above conditional input demand functions, we can write the cost relationship as

$$E = c(w,y,t)/\theta \tag{2}$$

where $E = \sum_j w_j x_j$ is the actual cost, and $c(w, y, t)$ is the minimum or frontier cost (cost without any inefficiency), defined as $c(w,y,t) = \sum_j w_j x_j(w,y,t)$. Sometimes, we define cost efficiency as the ratio of minimum to actual cost, i.e., $CE=(c, y, t)E= \theta \leq 1$. Since actual cost is increased due to technical inefficiency, $CE \leq 1$.

The estimation of a single equation stochastic cost frontier assumes the existence of data on the prices of the inputs employed, the quantities of outputs produced, and the total expenditure made by each of the I producers: Using the re-parameterization $\theta = \exp(-u),\ u \geq 0,$ we write the cost frontier in (2) as

$$\ln E_i = \ln c(y_i, w_i; \beta) + u_i \quad i = 1,2,...,I. \tag{3}$$

where β is the vector of technology parameters (that define the cost frontier) to be estimated, and $u_i \approx 1-\theta$ is input-oriented technical inefficiency. It can also be interpreted as the percentage increase in cost due to technical inefficiency. Note that in this formulation the input vector x_i used by the producer i need not be observed.

Given the above formulation, the cost efficiency (CE) of a producer i can be expressed as

$$CE_i = \frac{c(y_i, w_i; \beta)}{E_i} = \exp\{-u_i\}$$

which defines cost efficiency as the ratio of minimum possible cost to actual or observed cost. Since actual cost is greater than or equal to the minimum cost, it follows that the CE_i is always less than or equal to 1 and equals 1 only when the producer is efficient, i.e., actual cost equals minimum attainable cost.

In equation (3) the cost frontier $c(y_i, w_i; \beta)$ is deterministic because the entire excess of observed expenditure over minimum possible expenditure is assigned to cost inefficiency. However, sometimes cost increases can occur due to random exogenous shocks like weather, strikes, quality of inputs, etc., which are beyond the control of producers. In order to control for such exogenous factors, another random term is added to the cost function, and the model becomes :

$$\ln E_i = \ln c(y_i, w_i; \beta) + u_i + v_i \quad i = 1, 2, \ldots, I. \tag{4}$$

Under this formulation

$$E_i = c(y_i, w_i; \beta)\exp\{v_i + u_i\} \tag{5}$$

where $E_i = c(y_i, w_i; \beta)\exp\{v_i\}$ is the stochastic frontier. The stochastic frontier consists of two components : a deterministic part $c(y_i, w_i; \beta)$ that is common to all producers, and a producer-specific random part $\exp\{v_i\}$. We can calculate the producer specific efficiency exactly as before by :

$$CE_i = \frac{c(y_i, w_i; \beta)\exp\{v_i\}}{E_i} = \exp\{-u_i\} \tag{6}$$

and CE_i satisfies all the properties mentioned above.

Estimating equation (4) requires, (i) specification of a functional form for the deterministic kernel $c(y_i, w_i; \beta)$ (ii) an assumption about the distribution of the random variable v_i, and (iii) an assumption about the distribution of the random variable u_i. Given a particular specification, the Maximum Likelihood (ML) technique can be used to estimate the unknown parameters. Subsequently, the producer-specific inefficiencies can be estimated..Suppose we assume (the analytics of the ML technique does not depend on the particular specification for the deterministic kernel).

(i) $v_i \sim iidN(0, \sigma_v^2)$

(ii) $u_i \sim iidN^+(0, \sigma_u^2)$[1]

(iii) v_i and u_i are distributed independently of each other, and of the regressors.

Then the log-likelihood function for the sample of I producers can be written as (*see* Kumbhakar and Lovell (2000), pp. 140)

$$\ln L = constant - I \ln\sigma + \sum_i \ln\Phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right) - \frac{1}{2\sigma^2}\sum_i \varepsilon_i^2 \tag{7}$$

---

[1] Subsequent studies have assumed the more general specification of $u_i \sim iidN^+(\mu, \sigma_u^2)$ for the one-sided error component, and modeled as a function of other variables.

where $\varepsilon_i = u_i + v_i, \sigma^2 = (\sigma_u^2 + \sigma_v^2)$, $\lambda = \sigma_u / \sigma_v$ and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Battese and Corra (1977) suggested that the parameterization $\gamma = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}$ be used in place of $\lambda = \sigma_u / \sigma_v$, because γ has a value between zero and one, while λ could be any non-negative value. Thus the parameterization γ is better suited in obtaining the ML estimates as the search can be restricted within a known set. If we use the γ parameterization, the log-likelihood function is given by :

$$\ln L = constant - I \ln \sigma + \sum_i \ln[1 - \Phi(z_i)] - \frac{1}{2\sigma^2} \sum_i \varepsilon_i^2 \tag{8}$$

where $z_i = \frac{\varepsilon_i}{\sigma} \sqrt{\frac{\gamma}{1-\gamma}}$.

The log-likelihood function can then be maximized with respect to parameters to obtain ML estimate. Note that $\gamma(\lambda) \to 0$ when either $\sigma_u^2 \to 0$ or $\sigma_v^2 \to \infty$ and $\gamma(\lambda) \to 1$ when either $\sigma_u^2 \to \infty$ or $\sigma_v^2 \to 0$. In the first case, the stochastic frontier collapses to the OLS cost frontier (i.e., average response function) with no inefficiency, while in the latter case, the stochastic frontier collapses to the deterministic frontier with no noise. Thus, a log-likelihood ratio test on $\gamma(\lambda) = 0$ can be done to judge the appropriateness of the stochastic frontier analysis vis-a-vis the OLS cost function approach of modeling efficiency.

The estimated parameters can be used to obtain producer specific cost efficiency using the Battese and Coelli (1988) estimator, viz.,

$$CE_i = E(\exp\{-u_i\} | \varepsilon_i) = \left[\frac{1 - \Phi(\sigma_* - \mu_{*i} / \sigma_*)}{1 - \Phi(-\mu_{*i} / \sigma_*)} \right] \cdot \exp\{-\mu_{*i} + \tfrac{1}{2}\sigma_*^2\} \tag{9}$$

where $\mu_{*i} = \varepsilon_i \sigma_u^2 / \sigma^2$ and $\sigma_*^2 = \sigma_u^2 \sigma_v^2 / \sigma^2$.

Once we obtain estimates of cost efficiencies for each producers at each time period, two natural questions arise. The first question is, what is the behaviour of the inefficiencies over time? Are they increasing, decreasing or constant? Such a question is important especially in circumstances where policy interventions like deregulation, introduction of reforms, new entry, etc., as is the case in Indian banking, take place at particular points in time. One could look at the time behaviour of inefficiencies to judge the impact of these events on producer performance. The second question is, what explains the variations in inefficiencies among producers and across time. In particular, is inefficiency correlated with size, time, and ownership group? The second question encompasses the first, but looking at the former is often done to get an aggregative idea.

Following Kumbhakar (1990), Battese and Coelli (1992) proposed a simple model to estimate the time behaviour of inefficiencies. In their model, which was developed in a panel data context, the error term representing technical inefficiency was specified as :

$$u_{it} = \{\exp[-\eta(t-T)]\}u_i, \ i = 1,2,...,N; \ t = 1,2,...,T. \tag{10}$$

where the $u_{it} \sim N^+(\mu, \sigma^2)$, and η is a parameter to be estimated.

Under this specification, inefficiencies in periods prior to T depend on the parameter η. As $t \to T; u_{it}$ approaches u_T. Thus, inefficiency in period T can be viewed as the reference/benchmark point. Inefficiency prior to period T is the product of the terminal year's inefficiency and $\exp\{-\eta(t-T)\}$. If η is positive, then $\exp\{-\eta(t-T)\} = \exp\{\eta(T-t)\}$ is always greater than 1 and increases with the distance of the period t from the last period T. Thus when η is positive, inefficiencies fall over time. Conversely, when η is negative, inefficiencies increase over time. However, as Battese and Coelli note, the inefficiencies of defferent firms in any period t are equal to the same exponential function, $\exp\{-\eta(t-T)\}$, of the corresponding firm-specific inefficiency effects in the last period. Thus, the ordering of firms in terms of inefficiencies does not change during the period of analysis. This obviously rules out cases where relatively inefficient producers become more efficient over time and vice-versa. Though this is a restrictive feature of this model, we can interpret η as giving an "overall or average trend" in inefficiencies over the period of study across all producers, and thus this model provides a good starting point for modeling aggregative behaviour. Battese and Coelli implement this model in their FRONTIER program under the option "Model 1".

With respect to the second question, since the main motivation for efficiency analysis to policy makers is to design policies to improve performance of producers - especially the inefficient ones – it is highly desirable to know whether or not there are factors that can explain inefficiency. The variables explaining inefficiency are usually neither inputs nor outputs of the production process, but which nonetheless exert an influence on producer performance. They are thought to characterize the environment in which production takes place and so to influence the efficiency of production. Examples include the degree of competitiveness, input and output quality indicators; network characteristics, ownership form, changes in regulation, various management characteristics, and the like. To identify these factors, we use the model proposed by Battese and Coelli (1995), to determine the source of inefficiency. The Battese and Coelli model is similar in many respects to the models proposed by Huang and Liu (1994) and by Kumbhakar, Ghosh, and McGuckin (1991). Our preference for the use of the Battese and Coelli (1995) model is primarily because it is easily implementable since it is available as an option (Model 2) in the FRONTIER program. In brief, the model proposed by Battese and Coelli (1995), to explain production inefficiency within a panel data context, has the following structure :

$$\ln E_{it} = \ln c(w_{it}, y_{it}; \beta) + v_{it} + u_{it} \tag{15}$$

$$u_{it} = \delta' z_{it} + \varepsilon_{it} \tag{11}$$

where the first error term v_{it} represents the random noise in the production process of the *i*th firm in the *t*th period, and the second error term u_{it} captures the effect of technical inefficiency, which has a systematic component $\delta' z_{it}$ associated with the exogenous variables and a random component ε_{it}. The non-negativity requirement that $u_{it} = (\delta' z_{it} = \varepsilon_{it}) \geq 0$ is modeled as $\varepsilon_{it} \sim N(0, \sigma_t^2)$ with the distribution of ε_{it} being bounded below by the variable truncation point $-\delta' z_{it}$. Once the model is specified, the technological parameters and the inefficiency parameters are estimated by the MLE technique. Inefficiency estimates of individual producers at different time points are obtained as usual by the JLMS technique (*see* Battese and Coelli (1993) for the exact expresion of the log-likelihood function, and the FRONTIER manual (Coelli, 1996) for calculating the producer and time specific inefficiencies).

4. Data and Econometric Model

4.1 Data

Multiple outputs and multiple inputs characterize the banking industry. This is independent of whether a value added, a user cost, or an asset approach is used. However, a considerable disagreement exists in the banking literature on defining what exactly a bank produces. Following Berger and Humphery (1992), Grifell-Tatje and Lovell (1996) and Berg et al. (1993), we use the value added approach that treats both deposits and loans as outputs. Thus, our output vector *(y)* includes quantities (rupee value in 1980-81 prices) of deposits (*DEP*), loans and advances (*AD*), and investments (*IN*). Apart from these, we also include the number of branches (*BRN*) as an additional output. We do so keeping in mind the argument advanced in the literature that the number of branches can proxy for the quality and convenience that a bank offers to its customers (Grifell-Tatje and Lovell (1996), Berg et al. (1993)). Indeed, branch expansion, especially in the rural and semi-urban areas, has been an important objective of the regulatory policy of RBI. Labour (*L*) and capital (*K*) are the two variable inputs. Total cost is the total operating cost (*OPCOST*) of a bank, less costs that are of fixed nature like auditors' fees, lawyers' fees, etc. Price of labour (w_L) is obtained by dividing total expenses on labour by total number of employees. Similarly, price of capital (w_K) is obtained from w_K = (total operating cost – total expenses on labour) / total fixed assets.

The data for the present study is obtained from various issues of the reports (i) *Financial Analysts of Banks,* (ii) *Performence Highlights of Public Sector Banks,* (iii) *Performance Highlights of Private Sector Banks, and* (iv) *Performance Highlights of*

Banks published by the Indian Banks' Association. These publications report annual data from the profit and loss accounts and the balance sheets of all public and private banks operating in the Indian banking industry. The empirical model is estimated using data on 27 public banks and 23 private banks observed continuously over the years 1986 to 2000.[2] In 2000, these 50 banks accounted for 99.6 percent of all branches and 89 percent of total deposits of the commercial banking sector (Table 1).

4.2 The Econometric Model

We use a translog specification of the cost frontier to estimate the efficiency of the individual banks. The translog function has been widely used in efficiency studies and can be viewed as a second order approximation of any unknown cost function. The translog cost function in the present case is :

$$\begin{aligned}\ln E_{it} &= \alpha_0 + \sum_m \alpha_m \ln y_{mit} + \sum_j \beta_j \ln w_{jit} + \beta_t t \\ &+ \tfrac{1}{2}\left\{\sum_m \sum_l \alpha_{ml} \ln y_{mit} \ln y_{lit} + \sum_j \sum_k \beta_{jk} \ln w_{jit} \ln w_{kit} + \beta_{tt} t^2\right\} \\ &+ \sum_m \sum_j \alpha_{mj} \ln y_{mit} \ln w_{jit} \qquad (12) \\ &+ \sum_m \alpha_{mt} \ln y_{mit} t + \sum_j \beta_{jt} \ln w_{jit} t + u_{it} + v_{it}\end{aligned}$$

where $i = 1,....,I$ indexes banks and $t = 1,....,T$ indexes time.

We impose the usual symmetry restrictions on the above cost function, viz., $\alpha_{ml} = \alpha_{lm}$, and $\beta_{jk} = \beta_{kj}$. Moreover, to ensure linear homogeneity of the cost function in w, the following restrictions are imposed.

$$\sum_j \beta_j = 1, \sum_j \beta_{jk} = 0 \ \forall\ k, \sum_j \alpha_{mj} = 0 \ \forall\ m, \sum_j \beta_{jt} = 0 \qquad (13)$$

We estimate the above stochastic frontier models using the FRONTIER program developed by Coelli (1996). As outlined earlier, we estimate two alternative versions of the above model, namely the Battese and Coelli 1992 model for determining the time behaviour of efficiency, referred to as "Model 1" in the FRONTIER program, and the Battese and Coelli 1995 model for explaining inefficiency as a function of exogenous factors, referred to as "Model 2" in the FRONTIER program. Since our primary focus is on analizing the effect of ownership and deregulation on bank efficiency, we include one ownership dummy namely PVT (for private banks), one deregulation dummy, DEREG (which equals one if year > 1992; and zero otherwise), and a time variable t, and their interactions, while estimating Model 2. We estimate these models using only public and private banks that were in existence prior to the initiation of the reforms. Foreign banks are not included in the analysis because many of the regulations relating to the percentage of priority sector lending, branch expansion, etc., have been quite different for foreign banks compared to the domestic banks. Since efficiency is a relative concept, we preferred to keep the comparison set to be as homogenous as possible. For similar reasons, we did not include the new banks (the entrants) in our analysis.

---

[2] The year 2000 refers to the financial year beginning in April, 1999 and ending in March, 2000. Similarly, the year 1993 refers to the financial year beginning in April, 1992 and ending in March, 1993. We adopt this convention throughout the rest of the paper. Note that this convention is different than that in the paper by Kumbhakar and Sarkar (2003); and Sarkar, Sarkar and Bhaumik (1998).

The year 1993 is taken as the beginning of the post-liberalization period keeping in mind the fact that the first set of recomendations of the Narasimham Committee started being implemented in January 1992. The norms of income recognition, asset classification, and loan-loss provisioning changed quite substantially in the post-deregulation period due to the institution of prudential regulations and adoption of the BIS norms. The response to these changes has varied between bank groups,[3] as well as among banks within a particular group. Accordingly, in the econometric model that we estimate, the behaviour of inefficiency prior and post 1993 would be of significant importance.

5. The Alternative Models and Results

Table 2 presents some simple indicators of productivity of public and private banks. Deposits, advances and investments per employee of public banks were higher than those of private banks in 1985 and 1992. However, by 2000, this relative ranking had been reversed for each of the three indicators, with private banks showing a much higher growth rate (168 percent vis-a-vis 79 percent for deposits per employee, 153 percent vis-a-vis 51 percent for advances per employee and 231 percent vis-a-vis 130 percent for investments per employee) during the post-liberalization period. The situation with respect to cost per employee is just the reverse, with private banks exhibiting a slightly lower cost per employee than public banks in the pre-liberalization period, and slightly higher figures in the post-liberalization period. The average salary per employee of private banks were slightly lower than public banks in the pre-liberalization period, and comparable to public banks in the post-liberalization period.

Though Table 2 gives us an overall idea about the relative position of public and private banks in terms of some summary indicators, it is somewhat difficult to make judgments of efficiency using these indicators. In most situations where indicators of output are higher, so are the indicators of cost. To deduce efficiency propositions one needs a benchmark against which outputs and their associated costs can be judged.

---

[3] According to one estimate (MOF, 1993) the profits of the 28 public sector banks were reduced by 45 percent due to a switch to the new accounting norms.

Similarly, one trend is visible in the summary data, namely that there appears to be a structural break and a role reversal between public and private banks during the post-liberalization era. Whether this structural break and ownership difference is reflective of a general phenomenon or is driven by observations in a particular year or by a few banks needs to be evaluated in a statistical framework. This brings us to the estimation of the stochastic frontier models where costs are measured relative to the frontier level, and where efficiency can be explained in terms of exogenous variables like ownership and different time periods.

To conserve space we have not reported the parameter estimates, which for a translog cost function do not have any direct interpretations. However, we test for the presence of technical inefficiency in each model. Under the present formulation, testing for the presence of bank-specific inefficiency, and hence the necessity of using the frontier model, translates into testing the composite hypothesis $H_0 : \gamma = \mu = \eta = 0$.[4] The test is done using the usual likelihood ratio (LR) test, but the test statistic has a mixed chi-squared distribution (Battese and Coelli, 1996) and the critical value for a given level of significance is lower than the reported in the usual chi-squared tables. At the 1 percent level of significance, the critical value of the (usual) chi-squared distribution (with 3 degrees of freedom) is 11.341. The value of the test statistic in our case is much larger than this value (as noted above, the critical value is lower than 11.341) suggesting that our analysis overwhelmingly rejects the null hypothesis. Thus the standard average response functions not adequate for analyzing the cost behaviour of banks and a frontier model is required. Also note that the null hypothesis $H_0 : \gamma = 1$ is rejected at the 1% level of significance (the associated test statistic, which is asymptotically normally distributed, has value of –9.03), implying that a stochastic frontier specification fits the data better than a deterministic frontier. Thus the estimated parameters imply that the performance of banks are better analyzed within a stochastic frontier framework.

The estimated value of the parameter η is positive and significant at the 1% level. Recall that a positive value of η implies that ineffiencies of producers decrease over time. Thus our estimate suggests that the cost efficiency of Indian banks has improved during the estimation period. This is true for every bank since the parameter η in the Battese and Coelli (1992) model, is bank invariant. The calculated mean efficiencies[5] for each year are reported in Table 3 and represented in Figure 1. The means are reported for all banks as a group and separately for public and private banks. According to these estimates, the Indian banking system exhibits significant inefficiency, with the mean efficiency score varying from 69 percent in 1986 to 75 percent in 2000.It can also be seen that the mean efficiency score of public banks as a group is lower than that for private banks, the relative efficiency of the former being about 90 percent of the latter group. The mean efficiency of the banking system as a whole, as well as of each group, show an increasing trend; and in fact an exactly similar trend, with private banks being more efficient than public banks in every year.

---

[4] Note that in this model testing only the null hypothesis of $H_0 : \gamma = 0$ does not imply the absence of inefficiency as $\gamma = 0$ is consistent with the presence of bank-invariant inefficiency.

[5] Whenever we calculate mean efficiency scores, these are the simple means of the efficiency scores of the individual banks.

However, these observations could be reflective of the specification of the Battese and Coelli (1992) model. As outlined earlier, though the Battese and Coelli (1992) model is useful in obtaining an overview of the efficiency of the banking system, the model has two restrictive features, namely, (i) the time behaviour of efficiency is constrained to be a smooth monotonic function over the entire estimation period, and (ii) the ordering of banks is constrained to be the same for every year. Thus it is difficult to use this model in situations where exogenous shocks like reforms and deregulation are expected, and indeed targeted, to alter the operational efficiency of banks, as is indeed the case we are exploring. Also since different banks or bank groups can react differently to the reforms and deregulation measures, the rankings of banks in terms of their efficiency can be reasonably expected to change.

One way to adapt the Battese and Coelli (1992) model to handle such situations is to estimate the model separately for different time periods and different ownership groups (public and private). Assuming that one does estimate many such alternative models, comparisons of efficiency across periods and across groups are not strictly valid because one is allowing the benchmark (the frontier cost function) to change, which amounts to measuring efficiency changes with an elastic scale. The Battese and Coelli (1995) model that we outlined earlier overcomes many of these problems and is very well suited for analyzing the issues of the impact of deregulation and ownership, and it is to this model that we now focus our attention.

Four alternative models (translog cost function based on Battese and Coelli Model 2) are estimated for different specifications of the inefficiency function, i.e., the function used to explain bank inefficiencies in terms of exogenous variables (referred to as the Z variables). Like Model 1, the LR statistics for all the alternative models reconfirm that the stochastic frontier specification is the appropriate framework for analyzing bank performance; the LR statistic being significant at the 1 percent level for each of the four models.[6] Note that, the estimated value of γ is much lower in Model 2 than what we estimated in Model 1. This implies that the exogenous variables in the inefficiency function are able to explain a substantial part of the unconditional variance of the one sided error term.[7] We briefly note that the estimated coefficients of the translog cost function are theoretically consistent and about half of the twenty-eight parameters are significant at the 5 percent level in each of the four models.

---

[6] In Battese and Coelli Model 2, the null hypothesis of the absence of bank specific inefficiencies translates into $H_0 : \gamma = \delta_0 = \delta_1 .... \delta_m = 0$, where δ 's are the parameters associated with the Z variables.

We now examine the parameters of the inefficiency function reported in Table 4. Model A is the simplest, in which inefficiency is modeled in terms of a private ownership dummy. This model suggests, that if we look at the estimation period as a whole, then private banks as a group come out to be more efficient than public banks, a statistical confirmation of the result that we found based on the simple Battese and Coelli Model 1. Model B incorporates an additional time trend variable and allows this trend to be different for public and private banks. Both coefficients are negative, almost similar in magnitude, but are statistically insignificant. This suggests, that for the estimation period as a whole, we do not find any significant change in efficiency over time. Models C and D bring in the effect of deregulation into the analysis. Model C, is again a simple model, allowing only one time intercept shift in the inefficiency function in the deregulation period. The estimated parameter associated with the deregulation dummy is positive and significant, indicating that, on average (across years), the cost-efficiency of Indian banks has declined (i.e., inefficiency has increased) in the post-deregulation period. The coefficient associated with the interaction of the deregulation dummy and the private dummy is negative, suggesting that decline in efficiency of private banks has been less, but this coefficient is not significant even at the 20 percent level of significance.

Model D is the most comprehensive model. It allows ownership effects, time effects, and deregulation effects. In addition, the deregulation effects and the time effects are allowed to vary across ownership groups by incorporating suitable interaction terms. Looking at the pre-deregulation period one observes that the efficiency of banks has tended to increase over time, the coefficient associated with the time variable being negative and significant at the 5 percent level. Also, private banks appear to be more efficient than public banks. The effect of deregulation is captured in terms of the four variables (i) *dereg,* which gives the change in the intercept of the inefficiency function in the post-deregulation period, (ii) *dereg*t,* which gives the change in the slope of the inefficiency function with respect to the time variable, (iii) *dereg*pvt* and (iv) *dereg*t*pvt* which measure the differences in effects of *dereg* and *dereg*t*, respectively between public and private banks. We first observe that neither of the two variables *dereg*pvt* and *dereg*t*pvt* are significant at the 5 percent level (the associated p-values are 0.38 and 0.40 resapectively), implying that there are no ownership effects of deregulation and accordingly there are no significant changes in the relative efficiency of public and private banks in the post-liberalization period. The coefficient associated with the *dereg*t* variable is positive and significant suggesting that the efficiency of the Indian banks has exhibited a declining trend in the post-deregulation period *relative* to that in the pre-deregulation period. The coefficient is, however, of lower magnitude than that associated with the *t* variable. This implies, that the total effect (i.e., the sum of *t* and *dereg*t*) is still negative, so that the efficiency of Indian banks has continued to increase over the years in the post-deregulation period, albeit at a much slower rate compared to that in the pre-deregulation period. Finally, as observed earlier, since there are no significant ownership effects of deregulation on banking efficiency, and private banks were, on the average, found to be more efficient than public banks in the pre-deregulation period, the relative ranking has been preserved in the post-deregulation years.

---

[7] Refering to our discussion at the end of section 2, one can notice that in Battese & Coelli Model 2, the parameter $\gamma = \dfrac{\sigma_r^2}{\sigma_r^2 + \sigma_u^2}$.

Table 5 reports the annual average (over banks) efficiency scores for the banking system as a whole, as well as for public and private banks, implied by the inefficiency function based on Model D. Figure 2 gives a graphical representation of the figures in Table 5. A number of observations can be made from the table. First, in conformity with the estimated parameters for Model D, the efficiency scores for both public and private banks show an increasing trend from the year 1986 till the year 1992, the end of the pre-deregulation period. In the year 1993, there is a drop in efficiency of both bank groups, and since then a gradual increase. However, the rate of increase in efficiency over the years is less than the rate observed in the pre-deregulation period. By the year 2000, the last period of the sample, the efficiency level of the banking system as a whole, as well as of the public sector banks, were at the same level as in the year 1988, while for private banks the efficiency level was comparable to that in the year 1989. Efficiency levels in 2000, in general, were lower compared to the level reached at the end of the pre-deregulation period, namely, the year 1992. Second, the efficiency of private banks are higher than that of public banks, as a group, and this ranking is maintained during the entire period. Finally, the estimated average efficiencies based on the Battese and Coelli Model 2 are much higher compared to those from the Battese and Coelli Model 1. This is true for all models, A through D. This result is not surprising because Model 2 takes the determinants of inefficiency into account explicitly.

Table 6 presents the ranking of banks in terms of their average efficiency scores. Three rankings are given, one based on the entire estimation period (1986-2000), the next based on the pre-liberalization period (1986-1992), and finally the last based on the post-liberalization period (1993-2000). Under each of the three rankings, private banks appear to be more efficient than public banks, though to some extent this could be partly due to the parametric nature of the inefficiency function fitted in Model 2 (the private dummy being negative and significant). Some interesting observations can be made by looking at the change in ranking of the banks between the pre and the post deregulation periods. Private banks display a much wider variations in their ranks compared to that of public banks. While Spearman's rank corelation coefficient for public banks is 0.66, it is only 0.34 for private banks. Thus, within private banks, there are significant inter bank changes in rankings, with some banks leap-frogging over others in terms of relative efficiencies. The South Indian Bank Ltd., the Dhanalaksmi Bank Ltd., and the Lord Krishna Bank Ltd., exhibit substantial relative efficiency gains, while the City Union Bank Ltd., the Benaras State Bank Ltd., the Punjab Co-operative Bank Ltd., and the Bareilly Bank Ltd., exhibit significant loss in relative efficiencies. Compared to private banks, the rankings of public banks are much more stable over the two periods, with only the Oriental Bank and the Corporation Bank showing marked improvement in their rankings, and the Indian Bank showing significant decline in rankings.

The greater volatility of private banks in terms of their efficiency rankings between the pre and the post liberalization period is perhaps natural. The limits to branch and size expansion and other restrictions that existed prior to the reforms could have constrained private banks in a manner in which the efficiency differentials among banks could not materialize. This is particularly likely to be relevant when there are scale economies and previous studies (Kumbhakar and Sarkar, 2003) do indicate the presence of such scale economies in Indian banking. Under such circumstance, removal of constraints is likely to have different effects on different private banks. This explanation also fits quite well with the relatively lower volatility in the rankings of public banks, all of which were allowed to expand unfettered during the pre liberalization years.

6. Conclusion

In this paper we used stochastic frontier analysis to evaluate the efficiency of public and private sector banks in India over the period 1986 to 2000. The main focus was to examine if the liberalization measures of the 1990s had led to changes in banking efficiency and whether such changes varied across ownership groups.

Our results indicate that Indian Banks, on average, do exhibit the presence of cost inefficiency in their operations, though there is a tendency for inefficiencies to decline over time. The results indicate that recent deregulation has lead to an increase in the cost inefficiency of the Indian banks and a fall in the rate of inefficiency reduction. This phenomenon of increase in cost inefficiency subsequent to liberalization has been observed in many other banking studies as well, perhaps because liberalization brings with it significant changes in technology, procedures and practices, all of which shift the cost-frontier inwards, but individual banks are slow to respond and reorient themselves to these changes.

We also find the presence of some ownership effects; private banks, on average, are generally more cost-efficient than public banks. However, we do not find any significant differences in the impact of deregulation on the cost efficiency of these two ownership groups. At the individual level, we do find marked differences in the efficiency behaviour of different banks and their response to the deregulation measures Private banks show much more intra-group volatility in relative efficiency changes between the pre and post deregulation periods compared to that of public banks. This is perhaps because pre-deregulation constraints were more binding on private banks than on public banks. The ordering of the banks in terms of their ranking in the pre and the post deregulation periods can help us identify those banks which are in need of greater assistance and monitoring.

Finally, we make a few comments on the limitations of our analysis. First, in this analysis we have looked only at technical inefficiency. A bank could be both technically and allocatively inefficient. The cost function with both technical and allocative inefficiencies is more involved and difficult to estimate. Second, we have included branches as outputs. We have done so to capture the convenience and quality of service that banks offers to its customers. In our case, this inclusion has also been motivated by the regulator's objective of social banking so vigorously followed in India. However, it can be argued that including branches as outputs preempts the possibility of detecting inefficiency that can arise due to having a suboptimal number of branches. Since public sector banks in particular have a large number of branches, this could bias the (efficiency) estimates in favour of public banks. Therefore, examining the inefficiencies by excluding branches as outputs is worth exploring in the future. Finally, we have not considered directly the effect of new private banks as well as foreign banks on the efficiencies of the public and the old private banks, though we have incorporated the indirect effect as encapsulated in the proxy variable for time. Since efficiency is a relative concept, it would be instructive to include the new banks into the analysis to examine the effect of their entry on the relative standing of the public and old private banks.

References

Aigner, D. J., Lovell, C.A.K., and Schmidt, P. (1977) : 'Formulation and Estimation of Stochastic Frontier Production Function Models', *Journal of Econometrics, 6:1 (July), 21-37.*

Battese,G.E. and Coelli, T.J. (1995) : 'A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data', *Empirical Economics, 20, 325-32.*

Battese, G.E. and Coelli, T.J. (1993) : 'A Stochastic Frontier Production Function Incorporating a Model for Technical Inefficiency Effects', *Working Papers in*

Econometrics and Applied Statistics, No. 69, Department of Economics, University of New England, Armidale.

Battese, G.E. and Coelli, T.J. (1992) : 'Frontier Production Functions,Technical Efficiencies and Panel Data: With Application to Paddy Farmers in India', *Journal of Productivity Analysis 3.1/2 (June), 153-69.*

Battese, G.E. and Coelli, T.J. (1988) : 'Prediction of Firm-Level Technical Effiencies with a Generalized Frontier Production Function and Panel Data' *Journal of Econometrics, 38, 387-99.*

Battese, G.E and Corra G.S (1977) . 'Estimation of a Production Frontier Model: With Appliction to the Pastoral Zone off Eastern Australia', *Australian Journal of Agricultural Economics, 21:3,169-79.*

Berg, S.A., Forsund, F.R., Hjalmarsson, L. and Suominen, M. (1993) : 'Banking Efficiency in the Nordic Countries', *Journal of Banking and Finance, 17, 371-388.*

Berger, A.N. and Mester, L.J. (2001) : 'Explaining the Dramatic Changes in Performence of U.S. Banks : Technological Change, Deregulation and Dyanamic Changes in Competition', *Working Paper, University of Pennsylvania.*

Berger, A.N. and Humphrey, D.B. (1992) : 'Measurement and Efficiency Issues in Commercial Banking', *in Output Measurement in the Service Sectors, Studies in Income and Wealth, Z. Griliches (ed.), Vol. 56, The University of Chicago Press,Chicago, IL, for National Bureau of Economic Research, Chicago, 245-296.*

Bhattacharyya, A., Bhattacharyya, A. and Kumbhakar, S.C.(1997) : 'Changes in Economic Regime and Productivity Growth : A Study of Indian Public Sector Banks', *Journal of Compatative Economics, 25,196-219.*

Bhattacharayya, A., Lovell, C.A. Knox and Pankaj, S. (1996) : 'The Impact of Liberalization on the Productive Efficiency of Indian Commercial Banks', *European Journal of Operational Research, 21.*

Chaffai, M.E. (1997) : 'Productivity and Efficiency Performances of the Tunisian Banking Industry before and after the Economic Liberalization Program : An Economic Study Using Panel Data, *in Models for Economic Policy Evaluation Theory and Practice : An International Experiance, Research in Human Capital and Development, Dahel,R.and Sirageldin, I. (eds.), Vol. 11, Part B, JAI Press, 335-350.*

Coelli, T.J. (1996) : 'A Guide to FRONTIER Version 4.1 : A Computer Program for Stochastic Frontier Production and Cost Function Estimation', *Centre for Efficiencyand Productivity Analysis, University of New England, Armidale.*

Farell, M.J. (1957) : 'The Measurement of Productive Efficiency', *Journal of the Royal Statistical Society, Series A, General, 120, Part 3, 253-81.*

Grifell-Tatje, E. and Lovell, C.A.K. (1996) : 'Deregulation and Productivity Decline : The case of Spanish Savings Banks', *European Economic Review, 40, 1281-1303.*

Huang, C.J. and J.-T. Liu (1994) : 'Estimation of a Non-Neutral Stochastic Frontier Production Function', *Journal of Productivity Analysis, 5, 171-80.*

Indian Banks' Association : 'Performance Highlights of Public Sector Banks', Mumbai, India, Various Issues.

Indian Banks' Association : 'Performance Highlights of Private Sector Banks', Mumbai, India, Various Issues.

Indian Banks' Association : 'Performance Highlights of Banks', Mumbai, India, Various Issues.

Indian Banks' Association : 'Financial Analysis of Banks', Mumbai, India, Various Issues.

Kumbhakar, S. and Sarkar, S. (2003) : 'Deregulation, Ownership and Productivity Growth in the Banking Industry : Evidence from India', *Journal of Money Credit and Banking, 35.3 (June), 403-424.*

Kumbhakar, S.C., Lozano-Vivas, A., Lovell, C.A.K. and Hasan, I. (2001) : 'The Effects of Deregulation on the Performance of Financial Institutions : The Case of Spanish Savings Banks', *Journal of Money, Credit and Banking, 33, 101-120.*

Kumbhakar, S.C., and Lovell, C.A.K.(2000) : 'Stochastic Frontier Analysis', *Cambridge University Press, U.K.*

Kumbhakar, S.C., Ghosh, S. and McGuckin, J.T. (1991) : 'A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in US Dairy Farms', *Journal of Business and Economic Statistics, 9:3 (July), 279-86.*

Kumbhakar, S.C. (1990) : 'Production Frontiers, Panel Data and Time-Varying Technical Inefficiency', *Journal of Econometrics, 46,201-212.*

Mendes, V. and Reblo, J. (1999) : 'Productive Efficiency, Technological Change and Productivity in Portuguese Banking', *Applied Financial Economics, 9, 513-521.*

Ministry of Finance (MOF), Government of India : 'Public Sector Commercial Banks and Financial Sector Reform : Rebuilding for a Better Future', New Delhi : Govt. of India, 1993.

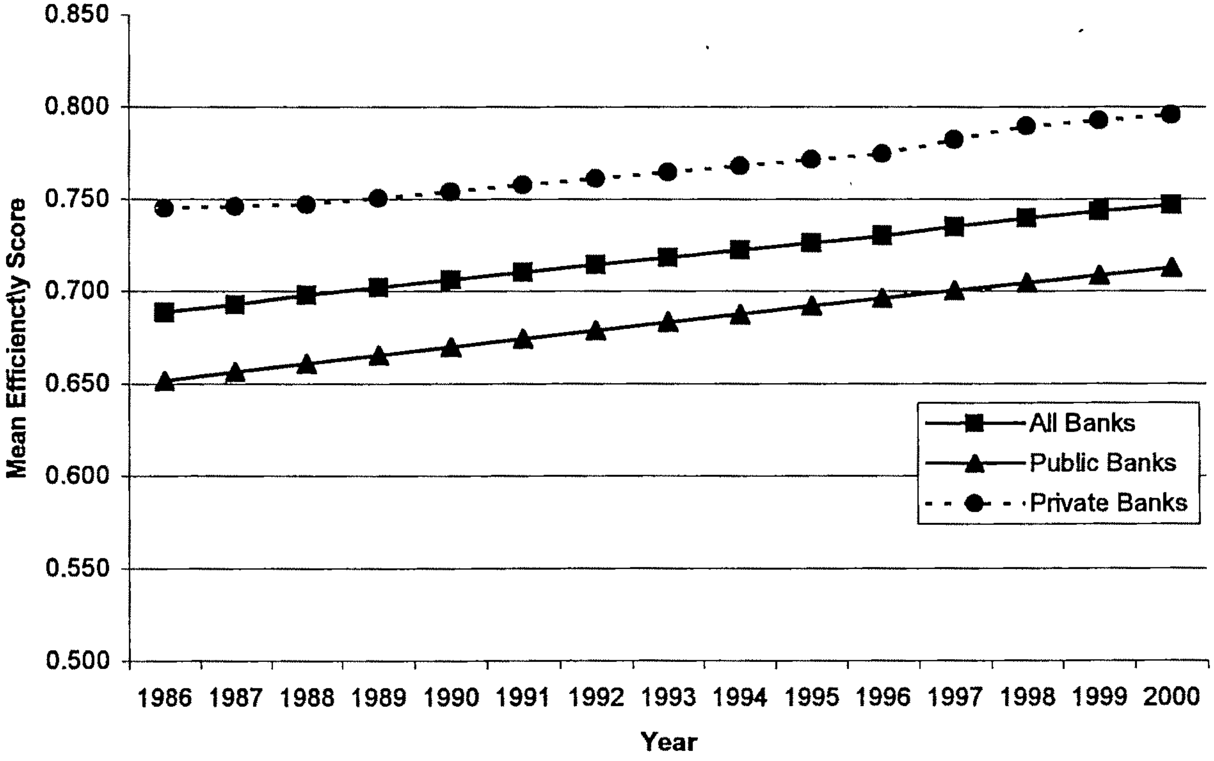
Meeusen, W. and Broeck, J. van den (1977) : 'Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error', *International Economic Review, 18:2 (June). 435-44.*

Sarkar, J. (1999) : 'India's Banking Sector : Current Status, Emerging Challenges and Policy Imperatives in a Globalized Environment', *in Financial Sector in India, James Hanson and Sanjay Kathuria (eds.), Oxford University Press.*

Sarkar, J., Sarkar, S., and Bhaumik, S. (1998) : 'Does Ownership Always Matter? Evidence from Indian Banking Industry', *Journal of Comparative Economics, 26, 262-281.*

Subramanyam, G. (1993) : 'Productivity Growth in India's Public Sector Banks : 1970-1989', *Journal of Quantitative Economics, 9, 209-223.*

Figure 1: Mean Efficiency of Indian Banks Based on the Simple Time Varying Efficiency Model



G165632

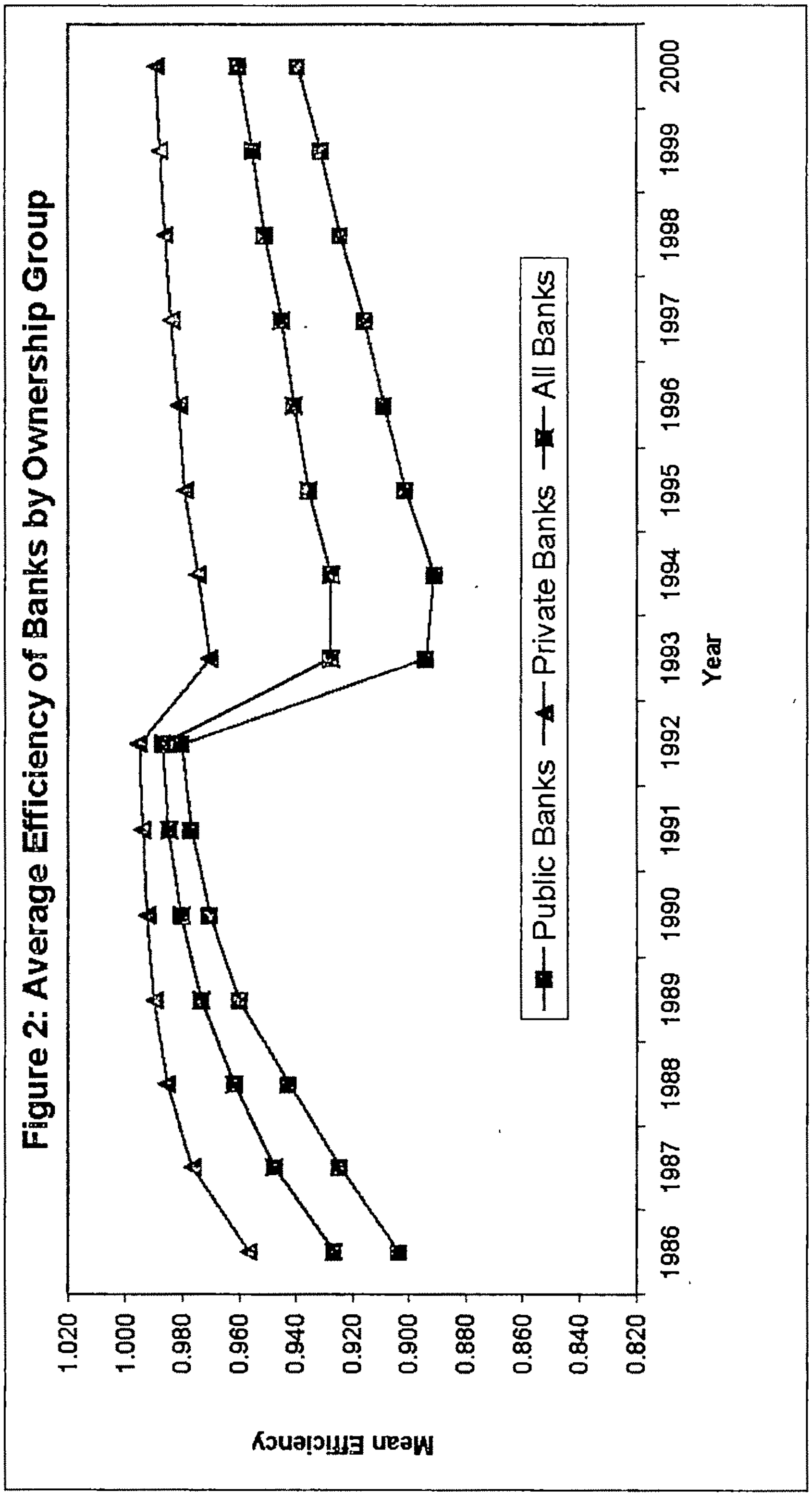


Figure 2: Average Efficiency of Banks by Ownership Group

Table 1: Institutional and Regulatory Structure of Indian Banking

	1986	1992	1997	2000
No of Banks	79	74	99	101
Public banks	28	28	27	27
Private banks	31	23	25	24
Foreign banks	20	23	23	22
New banks	-	-	24	28
				(20 foreign and 8 private)
Total Branches	40574	46396	49391	51540
Public banks	36161	42541	44752	46285
Private banks	4281	3717	4335	4606
Foreign banks	132	138	160	155
New banks	-	-	144	494
Total Deposits (Rs. Bn.)	1018.71	2649.18	5375.13	8688.42
Public banks	943.00	2359.20	4493.29	7060.44
Private banks	46.65	122.08	376.98	667.91
Foreign banks	29.06	167.89	367.69	478.95
New banks	-	-	137.17	481.12
Total Advances (Rs Bn.)	609.44	1590.50	2755.13	4434.66
Public banks	562.65	1436.46	2202.51	3521.10
Private banks	25.52	64.03	206.89	335.84
Foreign banks	21.26	90	257.44	338.20
New banks	-	-	88.29	239.52
Total Advances to priority sector (Rs. Bn.)	84.28	456.40	821.13	1306.63
Public banks	83.43	431.35	689.30	1082.07
Private banks	0.75	17.93	65.15	111.44
Foreign banks	0.10	7.11	44.75	71.67
New banks	-	-	21.93	41.45
Regulatory Features				
Cash Reserve Ratio (CRR) *		15.00	9.75	9.00
Statutory Liquidity Ratio (SLR)*		38.50	25.00	25.00
Lending Rates		Set by RBI	Free for loans > Rs. 200,000	
Deposit Rates		Set by RBI	Free	
Entry and branch expansion		Restricted	Free	
Accounting Norms		Companies' Act and RBI Directives	BIS Norms and RBI Directives	

All rupee figures are at nominal prices. Current exchange rate: 1 US $ = Rs. 48.39 (October 30th, 2002).
The Implicit GNP deflator (1980-81 = 100) was 150 in 1986, 259 in 1992 and 387 in 1997. CRR and SLR are as % of Net demand and time deposits.
Source: (i) "*India's Banking Sector: Current Status, Emerging Challenges and Policy Imperatives in a Globalized Environment*," by Jayati Sarkar in Financial Sector in India, James Hanson and Sanjay Kathuria (Eds.), Oxford University Press, 1999, and (ii) *Performance Highlights of Banks*, by Indian Banks Association, several issues.

Table 2: Simple Productivity Measures for Public and Private Banks

	1986	1992	2000
Public Banks			
Deposits per employee	2270.13	2994.42	5368.22
Advances per employee	1320.87	1714.37	2596.68
Investments per employee	677.41	1113.85	2560.05
Cost per employee	63.57	85.4	141.98
Wages per employee	54.29	70.84	118.85
Private Banks			
Deposits per employee	1826.74	2502.27	6717.18
Advances per employee	984.91	1302.94	3303.77
Investments per employee	524.11	833.88	2763.38
Cost per employee	60.74	79.92	150.11
Wages per employee	51.89	67.98	115.6

Note: All figures are in Rs. 1,000.

Table 3: Mean Efficiency Scores by Ownership Groups
(Based on Battese and Coelli, Model 1)

		Mean Efficiency	
Year	All Banks	Public Banks'	Private Banks
1986	0.689	0.652	0.745
1987	0.693	0.656	0.746
1988	0.698	0.661	0.747
1989	0.702	0.666	0.751
1990	0.706	0.670	0.754
1991	0.710	0.675	0.758
1992	0.715	0.679	0.761
1993	0.719	0.683	0.765
1994	0.723	0.688	0.768
1995	0.726	0.692	0.772
1996	0.730	0.696	0.775
1997	0.735	0.700	0.782
1998	0.739	0.705	0.790
1999	0.743	0.709	0.793
2000	0.747	0.713	0.796

Note: Sample period 1986-2000, All banks.

Table 4: Estimated Parameters of the Inefficiency Function Based on Battese and Coelli Model 2

	Model A	Model B	Model C	Model D
Intercept	**0.108**	**0.143**	**0.054**	**0.320**
Pvt	**-0.223**	**-0.155**	**-0.219**	**-0.151**
t		-0.006		**-0.060**
t*Pvt		-0.005		
Dereg			**0.103**	**-0.253**
Dereg*Pvt			-0.035	-0.321
Dereg*t				**0.051**
Dereg*t*Pvt				0.023

Bold numbers denote significant coefficients at the 5% level of significance

Pvt (private) and Dereg (deregulation), are dummy variables. Dereg=1 if year > 1992. The variable t stands for time, with t=1 for the year 1986. The interaction between the dummy variables time and private is captured by t*Pvt., and Dereg*t*Pvt is a three-way interaction.

Table 5: Mean Efficiency Score of Banks by Ownership Groups
(Based on Battese and Coelli, Model 2)

Year	All Banks	Public Banks	Private Banks
1986	0.926	0.903	0.956
1987	0.947	0.924	0.976
1988	0.962	0.942	0.985
1989	0.973	0.960	0.990
1990	0.980	0.970	0.992
1991	0.984	0.977	0.994
1992	0.987	0.980	0.995
1993	0.927	0.894	0.970
1994	0.928	0.891	0.975
1995	0.935	0.901	0.979
1996	0.941	0.909	0.981
1997	0.945	0.915	0.984
1998	0.951	0.924	0.986
1999	0.955	0.931	0.988
2000	0.960	0.939	0.989

Table 6: Ranking of Banks Based on Mean Efficiency Scores: Pre and Post Liberalization Periods (Battesse and Coelli , Model 2)

Bank	Group	Period 1986-2000	Period 1986-1992	Period 1993-2000
Jammu & Kashmir Bank Ltd	Private	1	2	5
Lord Krishna Bank Ltd	Private	2	5	1
City Union Bank Ltd.	Private	3	1	12
Karnataka Bank Ltd.	Private	4	6	4
Lakshmi Vilas Bank Ltd.	Private	5	3	6
Dhanalakshmi Bank Ltd.	Private	6	17	2
Nainital Bank Ltd	Private	7	7	11
Vyasa Bank Ltd.	Private	8	9	9
Bank of Madura Ltd.	Private	9	11	8
Benaras State Bank Ltd.	Private	10	4	15
Federal Bank Ltd	Private	11	12	7
South Indian Bank Ltd.	Private	12	21	3
Karur Vyasa Bank Ltd.	Private	13	16	10
Ratnakar Bank Ltd.	Private	14	14	13
Tamilnad Mercantile Bank Ltd	Private	15	15	14
Bharat Overseas Bank Ltd	Private	16	13	18
Catholic Syrian Bank Ltd.	Private	17	18	17
Bank of Rajasthan Ltd.	Private	18	20	16
Punjab Cooperative Bank Ltd.	Private	19	8	23
Sangli Bank Ltd.	Private	20	19	19
Bareilly Corporation Bank Ltd	Private	21	10	22
Nedungadi Bank Ltd.	Private	22	23	20
United Western bank Ltd.	Private	23	22	21
Allahabad Bank	Public	24	24	25
Dena Bank	Public	25	26	26
UBI	Public	26	28	28
Union Bank	Public	27	34	27
Oriental Bank	Public	28	39	24
Central Bank	Public	29	32	29
State Bank of Hyderabad	Public	30	30	31
State Bank of Patiala	Public	31	29	34
Andhra Bank	Public	32	33	32
Bank of Baroda	Public	33	31	33
Bank of Maharashtra	Public	34	27	36
Indian Bank	Public	35	25	39
Corporation Bank	Public	36	47	30
State Bank of Travancore	Public	37	42	35
Punjab & Sind Bank	Public	38	37	37
UCO Bank	Public	39	36	40
Punjab National Bank	Public	40	40	38
Indian Overseas Bank	Public	41	35	42
Vijaya Bank	Public	42	44	41
Bank of India	Public	43	43	43
State Bank of Indore	Public	44	41	46
State Bank of India	Public	45	38	44
State Bank of Mysore	Public	46	48	45
State Bank of Bikaner & Jaipur	Public	47	46	48
State Bank of Saurashtra	Public	48	50	47
Syndicate Bank	Public	49	45	50
Canara Bank	Public	50	49	49

Arthaniti 2 (1-2), 27-41 / 2003

Court Delays and the Defendant Behaviour : An Economic Analysis

Andrew Torre* and Bharat R. Hazari**

Abstract

Court delays are consistently criticised as being inimical to social welfare. However, the theoretical basis for this assertion is not well established in the law and economics literature. As a first step, very little is known about the impact of court delay on the defendant's optimal plea decision. If the defendant is rational in the sense of inter temporally optimising, court delay may increase or decrease the probability of a trial depending on the defendant's bail status. Some empirical support for this theoretical proposition is found using data on plea behaviour for a selection of cases heard in NSW Australia

JEL Classification : K41.

Keywords : Static cost function, Time dependant cost function, Optimal plea

1. Introduction

Court delays are a widely observed phenomenon. It is well documented that certain defendants have spent more time in jails awaiting a trial than what they would have spent if convicted for the original offence[1]. This particular set of events has occurred in several cases in India[2]. Trial delays are also quite common in many other countries, for example Australia, which are relatively rich compared to Third World Countries. These wealthier countries have the resources to expedite the trial process. There is a voluminous literature in law journals and books which analyse court delays from a legal perspective. Two hypotheses dominate this literature. Prior to the 1970's, court delays were attributed to large case loads and understaffed courts adopting poor management practices which an economist would view as an inefficient system. Current literature postulates that long waits are the outcome of the voluntary actions of key actors in the legal process such as lawyers[3]. Legal and popular thinking expresses various concerns about court delays[4]. First, it exacerbates the mental, social and financial burdens already borne by accused persons, who are presumed innocent until proven guilty. Second, it may reduce the defendant's probability of conviction due to deterioration in the quality of evidence. Third, it encourages guilty pleas and fourth it causes the erosion of public confidence in the judicial system, encouraging self-help as a remedy to address perceived wrongs[5]. While we recognize the importance of these legal and public concerns we will be concentrating on the economic analysis of trial delays from the perspective of the defendant who voluntarily chooses the plea strategy

---

* School of Economics, Deakin University

** School of Economics, Faculty of Business and Law, Deakin University, Pigdons Road

The object of this paper is to examine trial waiting time from the point of view of the defendants plea strategy. The prosecutor has been partially ignored from the analysis in the same spirit as is the separation in demand and supply analysis. In demand analysis the consumer maximizes utility subject to a budget constraint which takes the prices and income to be exogenously given. However, in a more sophisticated model the income variable is made endogenous by using a choice theoretic framework that incorporates labour-leisure choice. All these versions form building blocks to construct a general equilibrium model of an economy. An approach of going from simple to more complex frameworks does not invalidate the power and insight provided by ignoring some aspects of a particular problem. In this spirit our analysis concentrates solely on the defendant as an exceedingly important agent in the decision making process. The defendant's behaviour is analysed by using a decision-theoretic approach. The defendant is confronted by the prosecutor to make a choice between a guilty or a not guilty plea. Only in the latter case the defendant goes to trial while in the former case there is no need for a trial. Hence, the defendant has to choose between a guilty and not guilty plea bearing in mind that a not guilty plea involves waiting for a trial. Our analysis focuses on the impact of waiting time on the choice between a guilty and not guilty plea - this is analysed in terms of decision theory using cost functions and the expected value framework.

By using a decision–theoretic framework to analyse trial delays we arrive at several interesting results. First, we find out the probability p^* which makes the defendant indifferent between a guilty plea and the decision to wait and go to a trial. It is then established that this p^* is increasing in sentence and legal costs and decreasing in expected costs of sentence, costs associated with waiting for a trial and the legal costs of a trial. An *important* implication of this finding is that a reduction in the costs of a guilty plea results in more defendants taking up the option of pleading guilty and a reduced sentence. This reduces trial delays as those who accept guilty pleas move out of the queuing system. This, of course, reduces the social costs of trial delays. These results are based on a static framework. We then proceed to dynamize the static model and obtain another two important results. For the defendant not on bail, it is shown that when the actual waiting time is greater than the time at which the defendant is indifferent between going (not going) to trial then the defendant opts for the guilty plea. Similar results have been obtained for the on bail case. These results suggest that in countries (jurisdictions, crimes) which have long court delays the court system will be characterised by a relatively large volume of guilty pleas by not on bail defendants. Similar remarks can be made for the on bail defendants. The dynamics also highlight the difference between on bail and not on bail defendants. This behaviour is exceedingly important for policy purposes – specifically for designing policies that reduce court delays. We also provide empirical support for our hypothesis that defendants on bail opt for a trial as the delay increases. This is tested by using a logit model. The data used for this purpose is from the state of New South Wales in Australia and it does not reject the stated hypothesis regarding trial wait and defendant's behaviour in the on bail case.

2. The Defendant's Static Cost Function

We begin our analysis by constructing a cost function of the defendant in a static framework. The defendant faces the following choices: the sure option of a negotiated guilty plea with a discount in the punishment and a gamble of a not guilty plea which may result in a conviction or an acquittal. Let the defendant's costs of a guilty plea be denoted as:

$$ {}^dC_P^g = {}^dC_P^g[s,l] \quad \text{where} \quad \frac{\partial\,{}^dC_P^g}{\partial s} > 0 \quad \text{and} \quad \frac{\partial\,{}^dC_P^g}{\partial l} > 0 \qquad (1) $$

where s denotes sentence costs and l the legal costs of a guilty plea respectively. The function is increasing in both variables.

The expected cost function of the defendant for a not guilty plea is given below:

$$ E\,{}^dC_D^T = E\,{}^dC_D^T[PS,W,L] \qquad (2) $$

where the term PS denotes the expected costs of the sentence (as P denotes the probability of conviction)[5], W the waiting costs of a trial and L the legal costs of the trial. Note that the value of P lies in the closed interval $[0,1]$. If P equals one then the defendant is certain to be sentenced (similarly if P equals zero the defendant is certain to be not sentenced). The defendant faces the following choice: plead guilty, not face a trial and obtain some discount on the punishment; or plead not guilty and face a trial with an uncertain outcome. It is assumed that the defendant has to choose on the basis of the cost functions specified in equations (1) and (2). An examination of these functions shows that the function in equation (1) represents the certain option as it does not involve any random variables. The cost function for a trial has a random component denoted by PS. Hence, we have a situation in which the defendant has to choose between a certain and an uncertain option. Such a situation can be analysed by using expected utility theory and applying it to the defendant's problem.

Our first task is to identify a probability that will make the defendant indifferent between the cost of a guilty plea and the cost of going to a trial – in other words find P such that ${}^dC_P^g I E\,{}^dC_D^T$ (where I denotes indifferences). Formally, we require that:

$$ {}^dC_P^g(s,l) I E\,{}^dC_D^T = (PS,W,L) \qquad (3) $$

The left hand side of this is not a function of P but only s and l. As this is not a function of P it can be drawn as a straight line in the cost probability space as shown by line AB in Figure 1. The intercept A depends on s and l. The right hand side of equation (3) is a function of PS, W and L. Given W and L we can draw the line [illegible]

CALCUTTA ★ CENTRAL LIBRARY

which is a function of P. Note that it is drawn for a fixed value of W and L. The intercept C depends on W and L and is derived by setting $P=0$. The intersection of the two curves AB and CD at point e gives us, p^{*}, the probability at which the defendant becomes indifferent between pleading guilty and going to trial. From equations (1) to (3) it is clear that:

$$P^{*} = f(s,l;S,W,L) \tag{4}$$

The probability p' is an exceedingly important variable in our analysis. Let us determine its behaviour with respect to variations in s, l, S, W, and L. Changes in s and l shift the curve AB while changes in W and L shift the curve CD. A change in S pivots the curve. We arrive at the following proposition:

Proposition 1: The probability P^* (the probability at which the defendant becomes indifferent) is an increasing function of s and l and a decreasing function of $\bar{W}$, $\bar{L}$ and S.

This is demonstrated in Figure 1 by dashed lines. We will only provide an interpretation for one of these shifts. Let us suppose that s and l both increase, that is, the cost of sentence and the legal cost of a guilty plea. Such an increase shifts the curve AB to $A'B'$ and for a given CD curve p^{*} moves to p'^{*}, that is, the probability at which the defendant will be indifferent between a guilty plea and a not guilty plea increases. This probability has to rise as the cost gap between a sure choice and the not sure choice increases, and this implies that the P that makes the defendant indifferent must increase. Proposition 1 clearly shows the impact of costs on the probability of making the defendant indifferent between the guilty and not guilty pleas. An important implication of this comparative static result would be that a reduction in court delays could be reduced by lowering the cost of guilty pleas as more defendants would take up this option and move out of the queuing system.

Proposition 2: A reduction in the cost of guilty pleas results in more defendants taking up this option, hence, reducing trial delays by opting out of the system.

This is an exceedingly important result for policy makers who are concerned with the social costs of long delays in the delivery of justice.

3. Introducing Time in the Defendant's Cost Functions

In the previous section we have taken the important step of identifying the probability which makes the defendant indifferent between a guilty and not guilty plea. Equation (1) which shows the cost of a guilty plea is independent of time as it is assumed that guilty pleas are heard immediately by the court while a not guilty plea is time dependant (equation (2)). The time dependent cost function is given below:

$$EC_{DT}^{T} = EC_{Dt}^{T}\left[PS(t),W(t),L(t)\right] \tag{5}$$

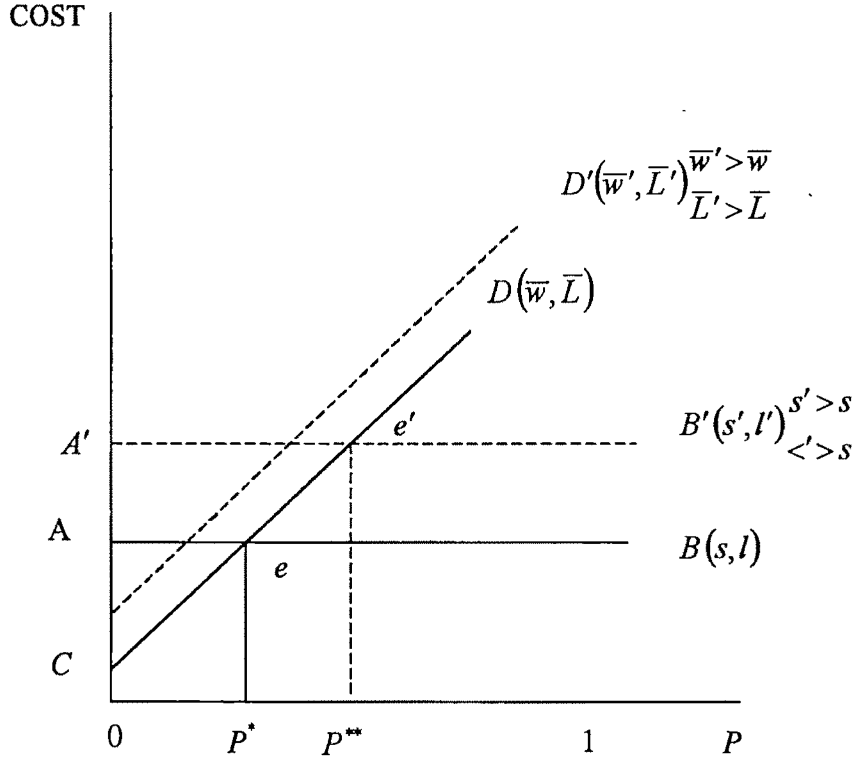


Figure 1

This is a general function which can be reduced to take account of the defendant who is on bail by setting $W(t)=0$, hence in the case of a defendant on bail equation (5) becomes:

$$EC_{Dt}^{T'} = EC_{Dt}^{T'}[PS(t), L(T)] \tag{5'}$$

The distinguishing feature of the above costs is that they are time dependant and require present value calculations. A specific formulation is given below:

$$S = \int_{t=0}^{T_1} ye^{-rt}dt + \alpha P\int_{t=T_1}^{T_1+L\lambda} ye^{-rt}dt - (1-\alpha P)\int_{t=T_1}^{T_1+L\lambda} ye^{-rt}dt \tag{6}$$

where:

T_1 = time at which trial begins

L = length of the sentence.

The above opportunity cost expression is (term by term): the present value of lost income due to being incarcerated while waiting for a trial; plus the present value of expected lost income due to being in jail if and when the defendant is convicted; minus the present value of income earned if the defendant is not convicted. Note that all costs are captured in y to keep the analysis as simple as possible.

It is easy to derive an expression for the defendant out on bail. It is a subset of the expression in equation (6) and is given below:

$$S' = \alpha P \int_{t=T_1}^{T_1+L} y e^{-rt} dt - (1-\alpha P) \int_{t=T_1}^{T_1+L} y e^{-rt} dt \tag{7}$$

The first term is absent as the defendant does not surrender any income while out on bail. By carrying out the integration in equations (6) and (7) we can obtain the cost functions for the defendant in both cases, i.e., not on bail (on bail). These are given below:

$$G = \frac{y}{r} e^{-rT_1} \left[(1-\alpha P)\left(e^{-rL} - 1\right) - \alpha P e^{-rL}\right] + \frac{y}{r} \tag{8}$$

and

$$G' = \frac{y}{r} e^{-rT_1} \left(1 - e^{-rL}\right)\left[2\alpha P - 1\right] \tag{9}$$

It is shown in the Appendix that G is increasing and G' decreasing in time provided that $2\alpha P - 1 > 0$. This is a surprising result as it shows that for those not on bail costs of waiting are falling over time. Hence they would make their choice in a different manner to those in jail waiting for a trial. It is important for policy makers to treat these groups differently to reduce trial delays and associated costs.

The point T_0^* which makes the defendant indifferent between going to trial and accepting plea bargaining is shown below. By differentiating G with respect to T and setting it equal to C_g^d and solving we obtain:

$$T_1^* = \frac{1}{r} \log \frac{\left[(1-\alpha P)\left(1 - e^{rL}\right) + \alpha P e^{-rL}\right]}{C_g^d} \tag{10}$$

It is clear that T_1^* is falling in C_g^d - so the cheaper the guilty plea is more defendants will accept it (can be easily seen from the diagram).

4. The Defendant's Optimal Solution in the Not On Bail Case

We now proceed to derive a decision making procedure for the defendant not on bail. The defendant has two choices. First, the defendant may be offered a reduction in punishment in exchange for a guilty plea. Second, the defendant has to wait for a trial. Given these two choices we demonstrate a procedure for arriving at a solution for the defendant in Figure 2. The horizontal axis shows time and the vertical axis costs as derived from the cost function G′. Let the curve EC denote the expected cost for the defendant derived for a given value of P* and r. The prosecutor's offer is shown by the line PP′. This shows the cost incurred by the defendant of accepting plea bargaining which reduces his sentence. The intersection of EC and PP′ at e provides a solution for the waiting time which makes the defendant indifferent between accepting the prosecutor's offer and going to trial. This is the point T_l^* derived in equation (10). Let $\overline{T}$ denote the actual trial wait. Since $T > T_l^*$ the defendant accepts the prosecutor's offer. Hence, we obtain the following proposition:

Proposition 3: The defendant not on bail accepts (rejects) the prosecutor's offer for $\overline{T} > T_l^* (\overline{T} < T_l^*)$.

Some comments are in order regarding the implications of Proposition 2. Let us assume that Figure 2 is based on a representative criminal, hence, it can be interpreted as an analysis of society as a whole. Proposition 2 then suggests that societies in which $T > T_l^*$ will be characterised by a large volume of guilty pleas. Similarly, for societies in which $T < T_1^*$ there will be relatively more trials. Note that this framework can be applied to each type of offence and for each court jurisdiction. For example, the analysis may be applied to sexual assaults, burglaries, murders and so on. If it is known that $T_1^* < \overline{T}$ then a guilty plea provides a way of reducing long waiting lines for a trial to occur.

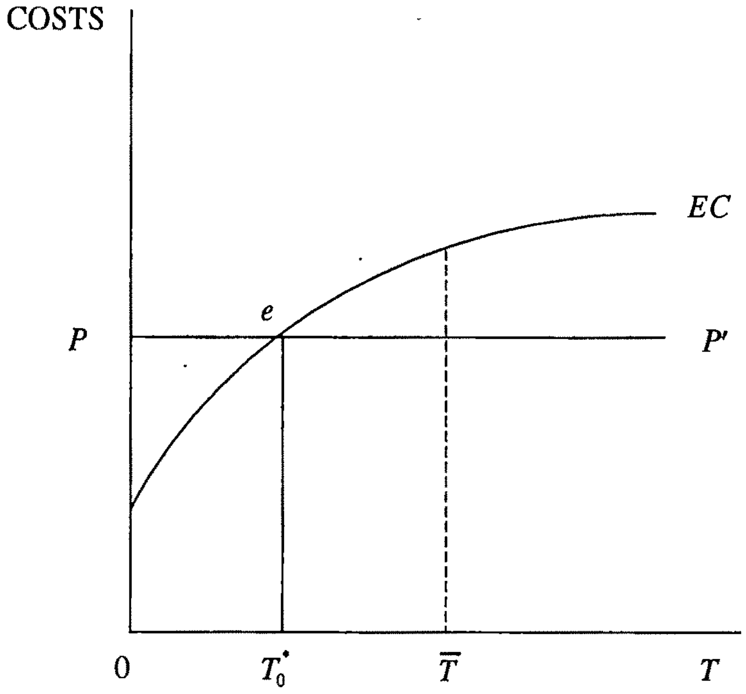


Figure 2

5. The Defendant's Optimal Solution in the On Bail Case

In discussing the on bail case it is important to first draw the expected cost curve. This is associated with equation (5′) which does not contain $W(t)$ by definition. As shown earlier this changes the shape of the cost curve which now becomes downward sloping. In Figure 3 the curve EC' shows the expected costs of the trial for the on bail case drawn for a given p' and r. The line pp' shows the prosecutor's offer giving rise to the indifference point T_0^*. If the actual trial wait $\overline{T} > T_0^*$ then the defendant goes to trial. The following proposition follows:

Proposition 4: The defendant on bail case rejects (accepts) the prosecutor's offer for $\overline{T} > T_0^* \left(\overline{T} < T_0^*\right)$.

Proposition 3 is the mirror image of Proposition 2. Comments made on Proposition 2 also apply to this Proposition with suitable adjustment towards its interpretation.

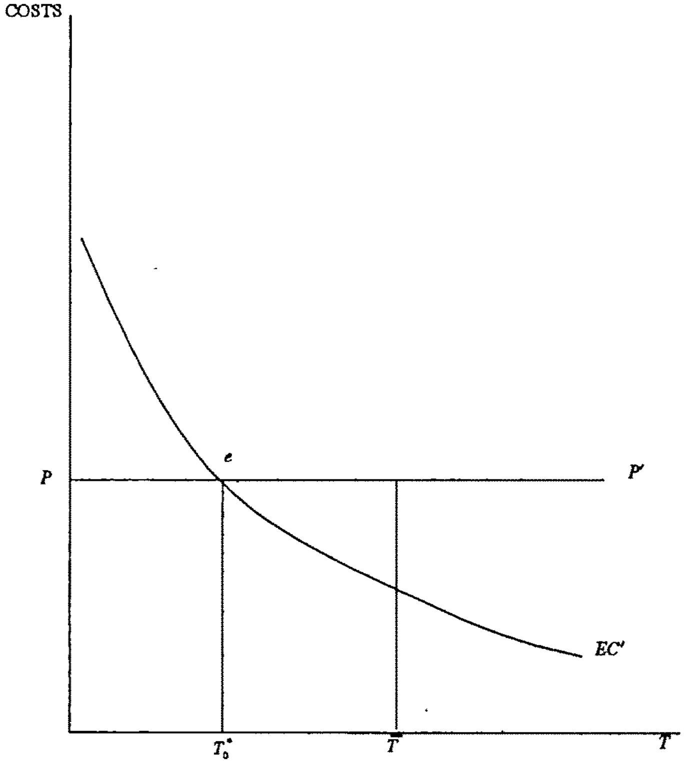


Figure 3

6. An Empirical Analysis of the Court Delays

In the previous sections of this paper we have established two Propositions. It is appropriate now to examine these Propositions empirically. We test these Propositions by using data from the criminal courts in New South Wales, Australia. We utilize a sequential logit model to explain plea behaviour and the factors which explain a defendant's acquittal and conviction after a trial and assume that each decision is independent of the other in order to simplify the resulting statistical analysis. This is a reasonable assumption to make, as at each decision stage different decision makers are involved. Conditional independence between the two decisions ensures that the cases that go to trial, as opposed to guilty pleas, constitute a random or representative sample of the set of all cases processed in the criminal courts.

A test of our model requires that we can observe each defendant's T^*, i.e., the point of indifference between going to trial and not going to trial. It is difficult to obtain information about T^* for each defendant. Nevertheless, our model suggests that we can proceed by postulating an equivalent hypothesis that states increases in the trial wait increases the probability of a guilty plea for a defendant not on bail. This can be seen from Figure 3. If $\overline{T}$ is increased then the probability of accepting a guilty plea increases as T_0^* is further away from the new T. A similar hypothesis can be derived for the defendant on bail.

Our model of the defendant's plea choice suggests that defendants not on bail will respond differently to increases in trial waiting time compared with defendants who are on bail. In this section of the study we use logit regression to empirically test this model for the on bail case. The data set is for defendants on bail, which comprises a sample of 363 defendants across six District Court areas in NSW for cases concluded in 1993, and two different offences, robbery and assault.

The estimating equations is given below:

$$Ln\left[\frac{P(D_i = 1)}{1 - P(D_i = 1)}\right] = \alpha_0 + \alpha_1 LAID + \alpha_2 PCONV + \alpha_3 EJAIL + \alpha_4 DELAY \tag{11}$$

This is estimated for each data set for the on bail case. The dependent variables used in the above equation are defined as follows; Dependent = 0, if defendant pleads guilty and 1 if the defendant went to trial and C is a constant. No court district or offence dummies are employed as average statistically significant effects suffice for the subsequent empirical work in this study. The remaining variables are defined below:

LAID = A binary variable which = 1 if the defendant was legally aided, = 0 otherwise

PCONV = A binary variable which = 1 if the defendant has prior convictions, = 0 otherwise

EJAIL = Mean expected jail sentence for offence in 1993

DELAY = Defendant's actual wait in years for trial

The legal aid variable is a proxy for legal costs in the plea choice model and requires no further comment. The defendant's prior conviction and bail status are used as proxy for sentence costs. Monetary and psychic imprisonment costs are proxied by the mean expected jail sentence in months handed down by the District Court in 1993 in a particular area and for a particular offence. We assume that defendants and their legal representatives make independent unbiased estimates of the likely jail sentence and its probability of occurrence which on average equal true mean expected jail sentence.

Since the imprisonment costs cannot be precisely observed and measured, particularly the psychic component, the logit regression results must be treated carefully. The results from estimating the logit regression for the bail case are given in Table 1. Results have also been obtained for the case in which the statistically insignificant variable has been dropped. These are presented in Table 2.

We now perform a likelihood ratio test to ascertain the statistical significance of the coefficients. The log of the likelihood for the unrestricted model is −129.0081 and for the restricted model −241.0846. The test statistic for the null hypothesis that the unrestricted model adds nothing to our knowledge about the determinants of the plea choice by defendants on bail charged with robbery and assault whose cases were completed in the NSW District Court in 1993 is 224.15. The critical chi-square value for the three degrees of freedom is 11.30 at the 0.01 probability level. Quite clearly the null hypothesis must be rejected so that the equation is statistically significant at the one percent level. The legal aid variable is not statistically significant in explaining plea choice, which stands in sharp contrast to popular perceptions in this matter. In general communities that have increased availability of legal aid are expected to be characterised by relatively more not guilty pleas by defendants. The hypothesis is not supported by our results. We now proceed to discuss other variables.

The predicted probabilities from the estimated equations are as follows. A defendant on bail with no prior convictions has a predicted probability of 0.69 of pleading not guilty and 0.31 of pleading guilty; which a similar defendant with prior convictions has a predicted probability of 0.39 of pleading not guilty and 0.62 of pleading guilty. If the trial was heard immediately the respective probabilities of going to trial would be 0.33 and 0.12 respectively. Consequently, it is clear that trial delays have a considerable impact on the decision to go to trial.

Conclusion

This paper by using a decision theoretic model analyses the behaviour of a defendant with regard to the impact of court delays. Two symmetrical propositions are established which relate to the behaviour of the defendants behaviours regarding the choice of a guilty plea or a trial. This choice is shown to depend on the expected cost function of the defendant, which is based on the expected value framework. The two propositions show that depending on the prosecutor's offer how the defendant decides whether to go to trial or not. It is shown that for the not on bail case the defendant accepts (rejects) the prosecutor's offer for same $\overline{T} > T_0^* \left(\overline{T} < T_0^*\right)$. This highlights the importance of time delays for the court system and defendant's decisions regarding trials. It shows that for societies with $\overline{T} > T^*$ there will be relatively more cases of acceptance of prosecutor's offer. Other propositions can be interpreted in a similar manner. We have also established that some of our theoretical results are not contradicted by the empirical results.

TABLE 1

Initial Logit Regression Results. Choice Of Plea By Defendants On Bail Charged With Robbery And Assault Whose Cases Were Concluded In The NSW District Court In 1993

Variable	Coefficient	Std. Error	T-Statistic	Prob
C	-5.255654	0.734180	-7.158536	0.0000
LAID	0.179021	0.326151	0.548888	0.5834
PCONV	-1.330764	0.325256	-4.091431	0.0001
DELAY	0.896609	0.229976	3.898710	0.0001
EJAIL	0.379718	0.052465	7.237598	0.0000

Log likelihood -128.8577
Obs with Dep = 1138
Obs with Dep = 0225

TABLE 2

Final Logit Regression Results. Choice Of Plea By Defendants On Bail Charged With Robbery And Assault In The NSW District Court In 1993

Variable	Coefficient	Std. Error	T-Statistic	Prob
C	-5.1817132	0.719533	-7.201491	0.0000
LAID	-1.308104	0.321895	-4.063765	0.0001
PCONV	0.895645	0.230069	3.892935	0.0001
DELAY	0.896609	0.229976	3.898710	0.0001
EJAIL	0.379718	0.052465	7.237598	0.0000

Log likelihood –128.8577
Obs with Dep =·1138
Obs with Dep = 0225

Variable	Mean All	Mean D = 1	Mean D = 0
C	1.0000	1.0000	1.0000
PCONV	0.490358	0.246377	0.64000
DELAY	1.681074	1.9572	1.511689
EJAIL	11.844	19.551	7.1177

Notes

1. Ms Kiran Bedi was given an award by UN for obtaining the release of several foreigners in India who had spent more time in jail than the original offence warranted as these alleged offenders could not obtain a hearing due to court delays.

2. The criminal court waiting time in Australia can be put into perspective by considering Table 1, which shows the median wait in days for defendants not on bail and on bail, from the defendant's arrest until the final outcome of the case, for all offences across all higher courts in NSW in 1992. Guilty plea dispositions are dealt with more quickly by the courts than trials, although for both disposition modes not

on bail defendants secure quicker hearings than their on bail counterparts for obvious reasons as shown in the following Table.

TABLE 1 Median Wait in Days For All Offences Concluded In The NSW Higher Criminal Courts In 1992

Bail Status			
Plea	On Bail	Not on Bail	Combined
NG_p	685 days	347.5 days	652 days
G_p	357 days	168 days	277 days

Source: New South Wales Criminal Courts Statistics 1992 NSW Bureau of Crime Statistics and Research

Where NG_p = Not Guilty and G_p = Guilty

As we can see from the Table above in the state of NSW the average waiting time for a trial is approximately 652 days in 1992. Of course, without proper analysis it is not possible to ascertain whether 652 days of waiting are optimal or otherwise.

As a first step in this direction we must understand how the critical decision makers, the prosecutor and the defendant reach their decisions. A recent NSW study on District Criminal Court time found that although more than 70 per cent of the criminal cases which go to hearing time consumed by guilty pleas is less than 5% of that of trials (NSW Bureau of Crime Statistics and Research 1992). Consequently, slight changes in the proportion of persons pleading not guilty are clearly capable of exerting major effects on the overall demand for District Court hearing time (NSW Bureau of Crime Statistics and Research 1992).

3. See Samuels (1991).

4. An idea regarding the extent of court delays can be found in the case of Australia from the documents of NSW courts.

5. For analytical convenience it is assured without loss of generality that P is independent of t. A relaxation of this assumption will not result in altering the main thrust of the analysis contained in this paper.

6. This is a standard procedure which is used in economic analysis of uncertainty. See any standard text book in Microeconomic Theory for example, Mas-Collel, A., Whinston, M.D. and Green, J.R., (1995).

Appendix

Shows the integration procedure and establishes the signs of the derivatives of the functions G and :

$$S = \int_{t=0}^{T_1} ye^{-rt}dt + \alpha P \int_{t=T_1}^{T_1+L} ye^{-rt}dt - (1-\alpha P)\int_{t=T_1}^{T_1+L} ye^{-rt}dt$$

Hence

$$G = -y\frac{e^{-rt}}{r}\bigg|_{t_1=0}^{T_1} - \frac{\alpha P y e^{-rt}}{r}\bigg|_{t=T_1}^{T_1+L} + (1-\alpha P)y\frac{e^{-rt}}{r}\bigg|_{t=T_1}^{T_1+L} \qquad \text{(A.1)}$$

$$= -\frac{ye^{-rT_1}}{r} + \frac{y}{r} - \frac{\alpha P y e^{-r(T_1+L)}}{r} + \frac{\alpha P y e^{-rT_1}}{r} + \frac{(1-\alpha P)ye^{-r(T_1+L)}}{r} - \frac{(1-\alpha P)ye^{-rT_1}}{r}$$

$$= -\frac{ye^{-rT_1}}{r} + \frac{y}{r} - \frac{\alpha P y e^{-rT_1}e^{-rL}}{r} + \frac{\alpha P y e^{-rT_1}}{r} + \frac{(1-\alpha P)ye^{-rT_1}e^{-rL}}{r} - \frac{(1-\alpha P)ye^{-rT_1}}{r}$$

$$= \frac{y}{r}e^{-rT_1}\left[-\alpha P e^{-rL} + (1-\alpha P)e^{-rL} - (1-\alpha P)\right] + \frac{y}{r}$$

$$= \frac{y}{r}e^{-rT_1}\left[(1-\alpha P)e^{-rL} - \alpha P e^{-rL} - (1-\alpha P)\right] + \frac{y}{r}$$

$$= \frac{y}{r}e^{-rT_1}\left[(1-\alpha P)(e^{-rL}-1) - \alpha P e^{-rL}\right] + \frac{y}{r}$$

It is easy to establish that this function is increasing in T_1.

$$\frac{\partial G}{\partial T_1} = -ye^{-rT_1}\left[(1-\alpha P)(e^{-rL}-1) - \alpha P e^{-rL}\right] > 0$$

as $\quad (1-\alpha P)(e^{-rL}-1) - \alpha P e^{-rL} < 0$.

For the defendant on bail the integrated function is

$$G' = -\frac{\alpha P y}{r}e^{-rT_1}\bigg|_{t=T_1}^{T_1+L} + (1-\alpha P)\frac{y}{r}e^{-rt}\bigg|_{t=T_1}^{T_1+L}$$

$$= -\frac{\alpha P y}{r}e^{-r(T_1+L)} + \frac{\alpha P y}{r}e^{-rT_1} + \frac{(1-\alpha P)y}{r}e^{-r(T_1+L)} - \frac{(1-\alpha P)y}{r}e^{-rT_1}$$

$$= \frac{y}{r}e^{-rT_1}\left[-\alpha P e^{-rL} + \alpha P + (1-\alpha P)e^{-rL} - (1-\alpha P)\right]$$

$$\frac{\partial G'}{\partial T_1} = -ye^{-rT_1}\left(1-e^{-rL}\right)\left[2\alpha P - 1\right] \text{ is negative as long as } \left(2\alpha P - 1\right) > 0 .$$

References

Carrington, P. (1969) : 'Crowded Dockets and The Courts of Appeals, The Threat to the Function of Review and the National Law', *Harvard Law Review, 82, 542-617.*

Maddala, G. (1983) : 'Limited Dependent and Qualitative Variables in Econometrics', *Cambridge University Press: Cambridge.*

Mas-Collel, A., Whinston, M. D. and Green, J. R. (1995) : 'Microeconomic Theory', *Oxford University Press: New York.*

Nimmer, R. (1978) : 'The Nature of System Change: Reforms in the Criminal Courts', *American Bar Foundation: Chicago.*

Rosenberg, M. (1965) : 'The Courts, The Public and the Law Explosion',*(Ed, Jones, H.) Prentice Hall, Englewood Cliffs N.J.*

Saari, D. (1967) : 'New Ideas for Trial Court Administration, Applying Social Sciences to Law', *Judicature, 51, 82-87.*

Samuels, G. (1991) : 'The Economics of Justice', *Journal of Judicial Administration, 1, 114-124.*

Weatherburn, D. and Nguyen Da Huong, M. (1992) : 'Aspects of Demand for District Criminal Court Time', *Crime and Justice Bulletin, no. 15 NSW Bureau of Crime Statistics and Research.*

Arthaniti 2 (1-2), 42-61 / 2003

Frontier Production Function and Extreme Observations

M. Pal, B. Ghosh and C. Neogi*

Abstract

Extreme Observations (EOs) play a significant role in the estimation of Frontier Production Functions (FPFs) The EOs can be defined to be the observations with either the maximum or the minimum efficiencies. The estimation of FPFs is known to be very much sensitive to the presence of EOs. It will be reviewed in this paper how the different models of single equation FPFs and the estimation procedures were developed in order to tackle the problem of EOs starting from the model proposed by Farrell (1957). The authors then propose a model of FPF with Errors-in-Variables (EIVs) and shows that the model may be effective even when the EIVs are negligible.

JEL Classification : C13.

Keywords : Frontier Production Function, Extreme Observations, Estimation and Errors-in-variables.

1. Introduction

The FPF gives the maximum attainable output for a given level of input. Though there is a vast literature on what should be the nature of such a function and on how this function should be estimated, the progress was initially hindered by the fact that a few, if not one, contaminated values may alter the picture altogether. Because the functions, after all, are to be estimated and the estimated values should be robust to any reasonably small change in the values of one or a few observations. The observations which may alter the values of the resulting estimates were identified and were termed as Extreme Observations (EOs); but there was no way to find out whether those values were subject to errors. Again the resulting estimates, besides being model sensitive, may also be sensitive to some extent to the particular method of estimations adopted for the purpose. Naturally some authors suggested a change in the model and others suggested a change in the estimation procedure.[1]

Before we define EOs, it is necessary to define efficiency. Farrell (1957) defined two kinds of efficiency of a firm: Technical Efficiency (TE) and Allocative Efficiency (AE). A firm is technically efficient if it is not possible for the firm to proportionately reduce its factors of production keeping the output level constant and the firm is allocatively efficient if it is not possible to reduce cost by reallocating the different levels of inputs keeping the output at the same level. Thus technical efficiency (TE)

---

* Economic Research Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700 108

[1] Readers are referred to
Forsund et al.(1980), Nishimizu & Page (1982), Bauer (1990), Sengupta (1990) Fried et al. (1993) among many others for comprehensive reviews and/or illuminating discussions on FPF. See also, Debreu (1951) and Farrell (1957) for pioneering works in this area.

may not imply allocative efficiency (AE). We shall throughout this paper discuss only TE, because we shall not bring price into consideration in our models.

Not all firms are technically efficient. EOs can be defined as the observations of the technically efficient firms. These are the first kind of EOs mostly discussed in the literature. The second kind of EOs are the observations with minimum efficiencies. These EOs are often outweighed by the first kind so far as the sensitivity of the resulting estimates are concerned. And thus most of the discussions will be based on the first kind of EOs. We shall of course give some passing remarks on the second kind of EOs whenever necessary.

Farrell assumed two factors of production with constant returns to scale production function to illustrate the concepts of TE and AE. Then he discussed the estimation through linear programming technique. Assuming constant returns to scale, the production function can be reduced to a single isoquant by a constant division of Y, the level of output, as

$$f(X_1, X_2) = 1$$

The plots of (X_1, X_2) will then look like the following diagram:

Diagram 1.1: Estimation of Efficient Production Function With Constant Returns to Scale and Two Factors of Production

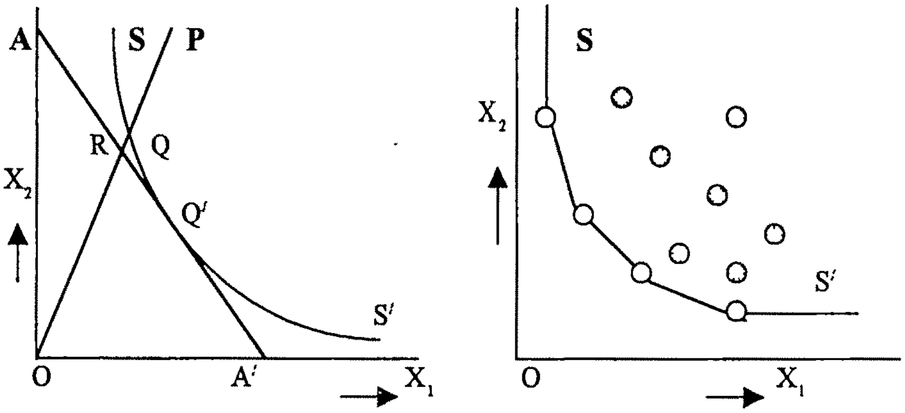


Diagram 1A
Technical and Allocative Efficiencies

Diagram 1B
Estimation of Efficient Production Function

(Constant Returns to Scale and Two Factors of Productions)

Naturally, SS′ is the most efficient production function which is also the isoquant with output level 1 unit. Given this isoquant, level of output for the most efficient firm can be found out very easily for any combination of inputs. The isoquant SS′, which can be estimated through linear programming technique, is obviously a convex function. All the points must lie on or above this function.

Generalizations to the case of many inputs and outputs are possible for this model. But it is not clear how constant returns to scale assumption can be relaxed for this particular model (See Kopp, 1981 for an attempt in this direction).

The linear programming models were later studied under the name of Data Envelopment Analysis (DEA) and received considerable attention from many authors.[2] The DEA models can tackle increasing returns to scale or variable returns to scale assumptions and are capable of answering many other questions. It is also possible to give models of production with many outputs. But these models or rather estimation procedures of these models are not robust to changes in the values of extreme observations.The model builders of DEA assume that the data do not contain any error. They also ignore a vital question which arises here. A firm may show technically inefficient due to climatic factors, misspecification of the functional form like non-inclusion of variables or wrong choice of the form itself. These factors have nothing to do with efficiency. Because efficiency is mainly viewed as the outcome of managerial ability.

It follows from the above discussions that the two limitations of the DEA models are Errors in Variables (EIVs) in the extreme observations which are generally assumed non-existent in the model, and a component of Errors in the Equation (EIQ) which is due to misspecification. The latter is the noise variable and does not depend on whether the firm is efficient or not (Aigner et al. (1977), Meeusen and Broeck (1977), Lee and Tyler (1978), Broeck et al. (1980), Stevenson (1980) etc.). The other part of the EIQ is due to technical inefficiency and is termed as inefficiency error. Assuming EIVs are non-existent, the observations may belong to either side of the frontier functions because of the presence of noise variable. All the observations must lie below the frontier when noise variables are absent, i.e., when the deterministic model is considered. (Aigner and Chu (1968), Timmer (1971), Afriat (1972), Richmond (1974), Schmidt (1976), Forsund and Jansen (1977), Forsund and Hjalmarsson (1979), and Greene (1980)). Due consideration should be given to the noise variable in the stochastic model while estimating the FPF. The error, other than the noise variable, termed as inefficiency error, when absent, gives rise to the usual regression model. Thus, existence of inefficiency error is a must for any FPF model to be meaningfully defined.

Once FPF is estimated, and the TEs are calculated for the firms, it is a usual task of the economists to see what makes the firms inefficient, if at all (Timmer (1971), Pitt and Lee (1981) and Cornwell et al. (1990)). For this they often bring other variables which are not included as inputs in the production function. These variables were primarily ignored as they were first thought to have insignificant additional influence on production or secondly thought to give rise to other complications such as autocorrelation, heteroscedasticity, multicolinearity, etc. Choosing appropriate functional form in the FPF model is perhaps as difficult as estimating it. The controversy centers on the question whether the form of FPF is same as that of average production function (APF). In stochastic framework the form of the two types of functions are assumed to be the same. The difference arises only from the assumptions on the two error terms – one taking all possible values on the real line (noise variable) and the other taking only non-positive values (inefficiency term). No plausible justification has so far been put forward in support of taking the form of FPF same as that of APF. Also various specific forms of distribution of inefficiency error terms have been assumed without giving proper reasons (Afriat (1972), Richmond (1974), Schmidt (1976), Aigner et al. (1977), Meeusen and Broeck (1977), Broeck et al. (1980), and Stevenson (1980)). Only in Lee (1983), Schmidt and Lin (1984) and Green (1980, 1990), there were attempts at testing the empirically found distribution of errors.

---

[2] For general reviews on Data Envelopment Analysis see Pettypool and Trout (1988), Seiford and Thrall (1990), Sengupta (1990)

The literature on this subject is vast. Since our aim is to discuss the development of the subject of FPF in order to tackle EOs, we shall mainly consider single equation FPF. The classification of econometric models as given by Forsund, Lovel and Schmidt (1980) will be followed throughout. In the next section we shall give these classifications.

In the subsequent sections we shall discuss the models as classified in the section 2. The estimation procedures and the sensitivity of the EOs will be highlighted. The advantage of using panel data models is pointed out in section 6. Section 7 presents EIV model, the estimation procedures of the model and shows why this model is better in view of the inevitable presence of EOs even if EIVs may be negligible.

2. Classifications of Econometric Models of Frontier

Let us start with the following equations:

$$\begin{aligned}\log Y &= \log f(X) - u + e \qquad (1)\\ &= \log f(X) + w\end{aligned}$$

or more specifically,

$$\begin{aligned}\log Y &= b_0 + \sum_1^k b_j \log X_j - u + e \qquad (2)\\ &= b_0 + \sum_1^k b_j \log X_j + w\end{aligned}$$

w is nothing but the EIQ, u is the inefficiency error and e is the noise variable. By definition u is non-negative.

A model is termed as nonparametric if no specific functional form of inputs is taken (i.e., equation (1)). Otherwise it is parametric (e.g., equation (2)). A model is termed as deterministic if the noise variable e is assumed to be absent in the model (i.e., equation (1) and (2) with e omitted). Otherwise it is stochastic. In addition there is another distinction. Explicit distributional assumptions are often made on the inefficiency error term u. The model may then be termed as statistical. Farrell's model is deterministic nonparametric (non statistical). Deterministic (non-statistical) models were studied later under the name DEA.

In the deterministic model, all the points must lie below the FPF. In other words, the FPF must pass through somewhere above all the points, because FPF is the maximum one can aim at. The maximum level of output that could be achieved for a given level of inputs is not known a priori for any of the production units. All we can do is to assume that some of the firms have achieved maximum technical efficiency, hence estimate the function and then compare other firms relative to it. This type of assumption makes the estimates data dependent, especially dependent on EOs of first kind, i.e., the observations which are thought to give hundred per cent or close to hundred per cent efficiency. This is what is meant in the literature when it is said that the estimated FPF is very much sensitive to EOs. The problem arises, because, the EOs are determined with respect to a plane, and the plane is estimated from the data which include the same EOs. It is not difficult to visualise that such EOs will always exist so long as we assume existence of some fully efficient production units.

3. Deterministic Parametric Frontier

Aigner and Chu (1968) and Timmer (1971) considered the following homogeneous Cobb-Douglas production frontier

$$\log Y = b_0 + \sum_{1}^{k} b_j \log X_j - u, \quad u \geq 0$$

u ≥ 0 forces all observations to be on or below the frontier. This model can accommodate non-constant returns to scale. This is described in the following diagram:

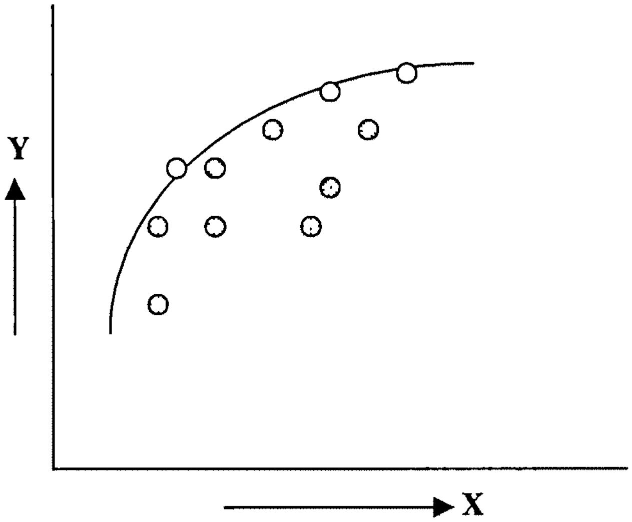
L_p-norm estimation methods drew maximum attention for this model. The method is to

$$\text{Minimize } \Sigma |u_i|^p$$

Subject to $\log Y \leq b_0 + \sum_{1}^{k} b_j \log X_j$ for all i.

This is a class of estimators which depends on p. Specific estimators can be found for specific values of p. For example, p=1 reduces it to a minimization of sum of absolute values of residuals and is known as Least Absolute (LA) Value (LAV) estimator. This is also known as LAE, LAR or LAD (E for Error, R for Residual and D for Deviation) estimator. This estimator can be reformulated so that it can be solved by Linear Programming (LP) method. It leads to Quadratic Programming method when

Diagram 2: Scatter diagram and FPF for single input: The deterministic case



p = 2 and in the limit i.e., when p tends to $\propto$, it leads to Chebyshev or Minimax method, in which case we minimize the maximum error.

The justification of taking specific values of p comes from the distributional assumptions on u. If the probability density function of u is

$$p(u) \sim \exp(-c|u|^{p})$$

then the Maximum Likelihood (ML) estimator is same as that of L_p-norm estimator. The corresponding distribution of u for p = 1, 2 and $\propto$ are respectively exponential, half-normal and uniform.

Needless to mention that the estimates are very much sensitive to EOs in this case. For small values of p the estimates are sensitive only to the EOs of first kind. As p increases the estimates become sensitive to both kinds of EOs. The first kind of EOs plays a dominant role because of the inequality restrictions. At p = $\propto$, we have to minimize the maximum error subject to inequality restrictions and thus EOs of second kind plays significant role here. Choice of the distribution of the error term is very critical. There is no apriori reason to take any specific distribution. The third criticism is that the range of dependent variable depends on the parameters to be estimated. This violates one of the regularity conditions which make MLEs consistent and asymptotically efficient.

The simplest way of overcoming the problem of EOs of first kind is to leave out a few EOs. There is still uncertainty involved in this method. How many EOs should we drop? Obviously the number of EOs to be dropped should be as small as possible. Moreover, one needs to check whether the estimates become stable at some stage. A more objective way of estimating the parameters is to use Corrected OLS (COLS) estimators. This is described in the next section. Another suggestion is to view the problem as arising due to EIVs. i.e., introduce EIVs in the model and estimate the parameters. This procedure is described in one of the subsequent sections. One can of course avoid taking this kind of model and consider stochastic frontier. Regarding choice of the distribution of errors, the work of Lee (1983), Schmidt and Lin (1984) and Greene (1980, 1990) deserve special mention. Tests of the appropriateness of the various distributional assumptions can be made using lagrange multiplier techniques as suggested by them. It has also been shown that the ML estimates are still consistent for gamma and half-normal density.

4. Corrected OLS Estimates

Richmond (1974) reformulated the model and wrote

$$\log Y = (b_0 - \mu) + \sum_1^k b_j \log X_j - (u - \mu); \qquad u \geq 0.$$

where now the error term $(u - \mu)$ (= e, say) has zero mean if $E(u) = \mu$. Now if we estimate $(b_0 - \mu)$ and the b_j's by OLS, it will automatically make the mean of residuals zero. So we cannot use the mean to estimate μ. If a specific distribution is assumed for u, then we can estimate μ consistently from the residuals using higher order moments. This estimate can then be used to correct the constant term to get corrected OLS (COLS) estimate. This is obviously consistent. For example,

(i) if u is one parameter gamma, $g(u) = (1/\Gamma(\sigma))\, u^{\sigma-1} \exp(-u)$, then $E(u) = V(u) = \sigma$, and $\hat{\mu} = \sum \hat{e}_i^2/n$.

(ii) if u is exponential, $g(u) = \exp(-u/\sigma)/\sigma$, then $E(u) = \sigma$, $V(u) = \sigma^2$ and $\hat{\mu} = \left(\sum \hat{e}_i^2/n\right)^{1/2}$ and

(iii) if u is half-normal, $g(u) = (1/\sigma).(2/\pi)^{1/2}. \exp(-u^2/(2\sigma^2))$. Then $E(u) = (2/\pi)^{1/2}\sigma$, $V(u) = \sigma^2. (\pi-2)/\pi$, and $\hat{\mu} = \left(\sum \hat{e}_i^2/n\right)^{1/2} \cdot \left(2/(\pi - 2)\right)^{1/2}$.

If the distribution of u involves two parameters, it is possible to estimate both using moments up to third order.

The problem with this method is that some of the residuals may have wrong signs. This may give problems in defining Technical Efficiency (TE). Moreover the correction depends heavily on the distribution assumed for u. Richmond (1974) himself got mean efficiency 87 per cent and 69 per cent by taking two different distributions of u. One can however correct the constant term so that no residual is positive and at least one is zero. This is also consistent (Gabrielson (1975), Greene (1980)). This is known as Modified COLS (MCOLS) or Modified OLS (MOLS) estimate. MOLS estimate, at least for the constant term, is again very much sensitive to EOs.

5. Stochastic Frontier Models (SFM)

Consider the following simple form of stochastic frontier model

$$\log \mathrm{Y} = \mathrm{b}_0 + \sum_{1}^{k} b_j \log \mathrm{X}_j - \mathrm{u} + \mathrm{v}, \qquad \mathrm{u} \geq 0.$$

$$= \mathrm{b}_0 + \sum_{1}^{k} b_j \log \mathrm{X}_j + \mathrm{e}, \quad \text{where } \mathrm{e} = \mathrm{v}\text{-}\mathrm{u}.$$

Existence of noise variable v guarantees the model to be a SFM. Appropriate estimation procedure would be to take either COLS estimates or ML estimates for this model. Both are consistent, but COLS is asymptotically less efficient than ML (Olsen et al. (1980)). However COLS outperforms ML in sample sizes of less than 400.

There is in the SFM no way of determining whether the observed performance is due to inefficiency or to random variation in the frontier, because it is not possible to get estimates of u and v separately. All we get is the estimate of e. Observation specific estimates of inefficiency can however be obtained by taking conditional expectation of u given e. Conditional mode can also be taken. E.g.,

$$\mathrm{E}(\text{-u} \mid e) = \mathrm{e}\sigma_u^2/(\sigma_u^2 + \sigma_v^2) - \sigma_u^2 \phi(A) / \left(\left\{\left(1 - \Phi(A)\right)\right\}\right)(\sigma_u^2 + \sigma_v^2)^{1/2}),$$

$$\text{where, A} = -\mathrm{e}/\{(\sigma_u^2 + \sigma_v^2)^{1/2}.\ \sigma_v/\sigma_u\}, \quad \text{and}$$

$$\mathrm{M}_0\,(\text{-u} \mid e) = \mathrm{e}\sigma_u^2/(\sigma_u^2 + \sigma_v^2) \qquad \text{if } \mathrm{e}<0$$

$$= 0 \qquad \text{if } \mathrm{e}>0$$

where $\mathrm{M}_0\,(\text{-u} \mid e)$ is the conditional mode of -u given e.

These estimates cannot be shown to be consistent estimates of u, since the variability of the conditional distribution u given e is independent of sample size. Since the variables are in logarithmic form, a better estimate of the inefficiency can be found from E(exp(u)|e). (Jondrow et al. (1982) and Kalirajan and Flinn (1983)).

6. Panel Data Models

Pitt and Lee (1981) first developed techniques to use panel data to estimate FPFs.[3] If panel data are available, one can estimate TE without specifying distributional assumptions on the error terms.

Schmidt and Sickles' (1984) model is

$$\mathrm{Y}_{it} = \alpha + \mathrm{X}_{it}\,\beta + \mathrm{u}_i + \mathrm{v}_{it}, \quad \mathrm{i} = 1,2,3,\ldots\ldots,\mathrm{N} \text{ and } \mathrm{t} = 1,2,\ldots\ldots,\mathrm{T}.$$

---

[3] See Pitt & Lee (1981), Schmidt & Sickles (1984), Battese & Coelli (1988, 1991), Cornwell et al (1990) and Kumbhakar (1990, 1991) among others for discussions about theoretical development and application of panel data.

Since u_i is time invariant, we can write

$$Y_{it} = \alpha_i + X_{it}\beta + u_i + v_{it}, \quad \text{where } \alpha_i = \alpha + u_i.$$

Estimates of the inefficiency disturbances can be obtained without assuming any particular distribution for these terms. Treat the most efficient firm in the sample as 100 per cent efficient (i.e., $u_i = 0$), so that the estimate of α is just the largest of the estimated α_i's.

Cornwell et al. (1990) allowed the levels of inefficiency to vary over time by firm replacing the firm effects u_i by

$$u_{it} = \theta_{i1} + \theta_{i2}t + \theta_{i3}t^2.$$

Kumbhakar (1990) started with a translog equation and proposed

$$u_{it} = \tau(t)\, u_i,\ t=1,2,\ldots\ldots,T, \quad \text{where } \tau(t) = \{1+\exp(bt+ct^2)\}^{-1},$$

where b and c are constants to be estimated. The resulting system must be estimated using MLE.

In Kumbhakar's model the relative efficiency is

$$\frac{u_{it}}{u_{jt}} = \frac{\tau(t)u_i}{\tau(t)u_j} = \frac{u_i}{u_j}.$$

This is independent of t. Thus the ranks of the efficiencies of the firms remain same over time. This is obviously a severe restriction. Because there are ample evidences that some firms which were less efficient than some other firms become more efficient than those other firms in course of time and vice versa. The model due to Cornwell et al. is a better model in this respect.

7. Errors-in-Variables and FPF

Economists while trying to tackle the problem of EOs, have not given much attention to, though not unaware of, one of the vital sources, namely the problems of Errors-in-Variables (EIVs). Timmer (1971, p.779) pointed out this problem and remarked, "Because only the extreme observations are used, the estimation of the frontier is highly subject to errors-in-the-data" (see also Forsund et al. (1980) and Bauer (1990) for more remarks). And then as a solution to the problem of first kind of EOs, he recommended the prevailing method of frontier to be "estimated in probabilistic fashion by constraining x percent of the observations to fall outside the frontier surface". This has been the practice of later economists until very recently.

This is nevertheless a good solution especially when nothing is known about the causes of errors leading to extreme observations. If it is already known that the errors-in-the-data exist, then a better way should be to incorporate it directly into the model and to solve the entire system. Here we give this EIV model and point out some interesting results. These results will explain why EIV model should be taken even if no EIVs are present.

In our model, the EIVs are assumed to be associated only with the regressors. Errors with the regressand if present do not pose much problem in estimation. The regression equation here,

$$Y_i = b_0 + b_1 X_{1i} + \ldots\ldots + b_k X_{ki} + \varepsilon_i,$$

for i = 1,2,...,n; contains regressors which are free from errors but are non-observable. The observed values are x_{ji} s such that

$$x_{ji} = X_{ji} + u_{ji},$$

for j= 1,2,...,k; i= 1,2,...,n. The subscript i refers to a production unit (usually a firm). X s are assumed to be non-stochastic. The EIVs are assumed to be independently distributed normal variates. They are also independent of the efficiency errors e. The firms take independent decisions so that they are independent among themselves.

More specifically,

$u_{ij} \sim N(0, \sigma_j^2)$

and $\varepsilon_i \sim$ half-N(0, σ^2) for all i.

The likelihood function can be written as

$$L' \propto (1/\alpha) e^{-(T_i^2/2\sigma^2)} \prod_{j=1}^{k} (1/\sigma_j) e^{-(X_{ji}-x_{ji})^2/(2\sigma_j^2)}$$

where, $T_i = Y_i - b_0 - b_1 X_{1i} - \ldots\ldots - b_k X_{ki}$.

This involves too many unobserved variables. The best way to do away with these variables is to integrate these variables out subject to the condition $T_i \leq 0$. Because of the assumption of normality of EIVs the likelihood function then becomes

$$L' \propto \prod_{i=1}^{n} \int_{-\infty}^{0} (1/\alpha) e^{-T_i^2/(2\sigma^2)} . (1/b) e^{-(T_i-a_i)^2/(2b^2)} dT_i .$$

Where, $a_i = Y_i - \beta_0 - \beta_1 X_{1i} - \ldots\ldots - \beta_k X_{ki}$,

and $b^2 = \beta_1^2 \sigma_1^2 + \ldots + \beta_k^2 \sigma_k^2$.

After simplifying the expression further we get

$$L' \infty \prod_{i=1}^{n} (b^2 + \sigma^2)^{-1/2} e^{-a_i^2/\{2(b^2+\sigma^2)\}} \{\Gamma_\infty(1/2) + \Gamma_{d_i}(1/2\})\}$$

where $d_i = a_i^2 \sigma^2 / \{2b^2(b^2+\sigma^2)\}$,

and $\Gamma_d(p) = \int_0^d t^{p-1} e^{-t} dt$.

$\Gamma_\infty(p)$ is nothing but the usual gamma function and is usually written as $\Gamma(p)$. In our case, in particular, $\Gamma_\infty(1/2) = \Gamma(1/2) = \pi^{1/2}$.

The parameters which are to be estimated from the likelihood function are the regression parameters and the error variances. a_i's and hence d_i's are functions of unknown regression parameters. The variances of the EIVs always appear in the form of b^2. Thus we need not know each of the error variances separately to maximise the likelihood function. It is necessary to write b^2 as $\sigma^2(b^2/\sigma^2)$ or $\sigma^2\,\sigma_0^2$ (say), so that the part of b^2 other than σ^2 (i.e., σ_0^2) becomes invariant under change of unit of measurement. The likelihood function is then

$$L' \propto \prod_{i=1}^{n} \{\sigma^2(1+\sigma_0^2)\}^{1/2} e^{-a_i^2/\{2\sigma^2(1+\sigma_0^2)\}} \{\Gamma(1/2) + \Gamma_{d_i}(1/2)\}$$

where, $d_i = a_i^2/\{2\sigma^2\sigma_0^2(1+\sigma_0^2)\}$

We maximise the log-likelihood function which can be written apart from the constant term as

$$\text{Ln.L}' \propto -(n/2)\ \text{ln.}\sigma^2 - (n/2)\ \text{ln.}(1+\sigma_0^2) - \Sigma\, a_i^2/\{2\sigma^2(1+\sigma_0^2)\} + \Sigma\ \text{ln.}\{\Gamma(1/2) + \Gamma_{d_i}(1/2)\}.$$

To maximise this function, one can further reparametrise σ^2 as

$$\sigma_*^2 = \sigma^2(1+\sigma_0^2).$$

Now, it is easy to visualise the behaviour of σ_0^2, if we view the log-likelihood function as a function of σ_0^2 and σ_*^2 instead of σ_0^2 and σ^2. Ln.L' is then a decreasing function of σ_0^2, the maximum value attaining at $\sigma_0^2 = 0$. This is due to the identification problem usually faced in the EIV model.

The EIV model described above is not completely identified even otherwise. There exist infinitely many combinations of $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ which give the same value of b^2. One way to make the frontier model identified is, as in the usual model, to assume $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ to be known apart from the constant multiplier of σ^2. Symbolically, we assume

$$\sigma_j^2 = c_j^2\sigma^2, \qquad \text{for all } j$$

and, c_j^2 values should be known and supplied. It is, however, not necessary to know these values so long as we are interested only in estimating regression coefficients. The knowledge on the value of σ_0^2 serves the purpose. Thus we maximise the likelihood function either for known values of c_j^2 s or for a given known value of b^2, since b^2 involves other unknown regression parameters as well.

There is even another problem with this type of estimation procedure. It is not possible to estimate the individual errors $\hat{\varepsilon}_i$'s, since u_{ji} s are being integrated out. All we can do is to get the expected values of $\hat{\varepsilon}$ given the observed values, which is nothing but $\hat{a}_1$. Calculation of technical efficiencies then becomes straightforward. This problem, however, exists in all kinds of stochastic FPF model (Forsund, Lovell & Schmidt (1980)).

7.1 An Illustration

Data on output (value added) and inputs (labour and capital) for 29 modern and 29 traditional industries over 15 years were collected from various sources as described in Pal, Neogi and Ghosh (1996). The following simplest Cobb-Douglas form

$$V = AK^{\alpha} L^{\beta}$$

was fitted after taking logarithmic transformation. The estimates were found separately for $\sigma_0^2 = 0$ (i.e., MOLS estimates), $\sigma_0^2 = .01$, $\sigma_0^2 = .05$ and $\sigma_0^2 = 0.1$. The efficiencies of the industries were calculated and then ranks of the efficiencies were obtained for each year from 1974-75 to 1988-89. Table 1 shows the estimated regression coefficients for $\sigma_0^2 = 0.05$.

Table I : Estimated Coefficients of The Frontier Production Function

(at 5% EIVs)

Year	Intercept	α	β	σ_0^2	% above FPF
Modern Industries					
1974-75	4.04	0.55	0.53	0.40	6.90
1975-76	4.50	0.45	0.71	0.45	6.90
1976-77	4.61	0.41	0.84	0.31	6.90
1977-78	5.17	0.35	0.89	0.29	6.90
1978-79	3.86	0.50	0.74	0.40	6.90
1979-80	4.39	0.44	0.76	0.27	6.90
1980-81	5.39	0.31	0.97	0.36	10.34
1981-82	6.31	0.22	1.06	0.34	6.90
1982-83	6.07	0.20	1.20	0.31	10.34
1983-84	4.48	0.49	0.63	0.42	6.90
1984-85	5.39	0.29	1.09	0.25	10.34
1985-86	5.87	0.20	1.29	0.33	6.90
1986-87	6.24	0.20	1.20	0.41	10.34
1987-88	4.50	0.35	1.11	0.43	10.34
1988-89	4.87	0.32	1.13	0.54	10.34

Year	Intercept	α	β	σ_0^2	% above FPF
Traditional Industries					
1974-75	3.99	0.52	0.52	0.38	10.34
1975-76	4.35	0.47	0.57	0.38	10.34
1976-77	4.38	0.49	0.54	0.51	10.34
1977-78	3.82	0.52	0.52	0.36	10.34
1978-79	4.46	0.47	0.55	0.55	10.34
1979-80	4.70	0.44	0.61	0.51	6.90
1980-81	4.91	0.42	0.61	0.68	6.90
1981-82	4.68	0.44	0.55	0.42	6.90
1982-83	5.11	0.40	0.60	0.43	3.45
1983-84	4.39	0.45	0.59	0.40	3.45
1984-85	6.78	0.22	0.81	0.51	6.90
1985-86	9.12	0.02	1.02	0.84	6.90
1986 -87	4.88	0.45	0.53	0.61	3.45
1987-88	8.25	0.13	0.82	0.54	6.90
1988-89	7.27	0.21	0.81	0.57	10.34

Table 2 shows how the coefficients changed as the extent of errors was assumed to increase selectively for some years. The coefficients more or less stabilize as σ_0^2 increases. This phenomenon becomes more clear if we look at Table 3. Table 3 shows the rank differences between the pair of cases (i.e., the difference of rank of efficiency of an industry obtained for some value of σ_0^2 to the rank of the efficiency of the same industry for some other value of σ_0^2) selectively for the years 1974-75, 1981-82 and 1988-89. The picture is similar for other years. Differences of ranks of efficiencies were shown for the following pair of cases: (MCOLS vs. $\sigma_0^2 = .01$), ($\sigma_0^2 = .01$ vs. $\sigma_0^2 = .05$) and ($\sigma_0^2 = .05$ vs. $\sigma_0^2 = .10$). Observe that σ_0^2 is nothing but the ratio of variances of EIVs over that of inefficiency error. The rank differences are large initially and then these almost vanish at the end. Thus we can say that the coefficients and especially the ranks of efficiencies stabilize at $\sigma_0^2 = 0.05$ if not prior to that. This indicates that it may be sufficient to assume 5% EIVs even if the actual extent of EIVs is much higher/ lower.

It is worth mentioning here that if we do not take EIV model the MCOLS estimates differ much if one or more extreme observations are omitted from the set of data.

Neglecting x percent of extreme observations of first kind and introducing such type of errors in the data are, however, not equivalent. It is rather equivalent to shifting some or all of the x_{1i}, x_{2i}, ... , x_{ki} and y_i values to the desired direction for the firms concerned. If we stick to the principle of minimal shift, this will imply shifting only in one direction so that u_{jt}, for a given j, will always either be non-negative or be non-positive; whereas EIV model assumes that the average/expected errors must vanish.

Table 2: Estimated Regression Coefficients of the FPF For Different Values of Relative Variances of EIVs.

Year	MOLS		$\sigma_0^2 = .01$		$\sigma_0^2 = .05$		$\sigma_0^2 = 10$	
	α	β	α	β	α	β	α	β
			Modern Industries					
1974-75:	.48	.65	.69	.22	.55	.53	.51	.62
1979-80:	.38	.89	.41	.98	.44	.76	.42	.80
1984-85:	.37	1.00	.27	1.11	.29	1.09	.30	1.09
1988-89:	.45	.86	.35	1.10	.32	1.12	.34	1.11
			Traditional Industries					
1974-75	.60	.43	.53	.49	.52	.52	.53	.52
1979-80	.52	.55	.42	.60	.44	.61	.45	.61
1984-85	.27	.78	.28	.69	.22	.81	.22	.82
1988-89	.34	.70	.14	.89	.21	.81	.26	.77

Table 3: Number of cases of change in Efficiency Rankings Due to Change in Variances of EIVs

	Change		**Rank Differences (No. of Cases)**			
Year	**From**	**To**	**2**	**3**	**4**	**5 or more**
		Modern Industries				
1974-75	MCOLS	$\sigma_0^2 = .01$	3	3	3	6
	$\sigma_0^2 = .01$	$\sigma_0^2 = .05$	6	2	2	4
	$\sigma_0^2 = .05$	$\sigma_0^2 = .10$	2	0	1	1
1981-82	MCOLS	$\sigma_0^2 = .01$	4	1	2	5
	$\sigma_0^2 = .01$	$\sigma_0^2 = .05$	3	4	1	1
	$\sigma_0^2 = .05$	$\sigma_0^2 = .10$	0	1	0	0
1988-89	MCOLS	$\sigma_0^2 = .01$	5	2	2	0

	$\sigma_0^2 = .01$	$\sigma_0^2 = .05$	1	0	0	0
	$\sigma_0^2 = .05$	$\sigma_0^2 = .10$	1	0	0	0
		Traditional Industries				
1974-75	MCOLS	$\sigma_0^2 = .01$	2	2	1	1
	$\sigma_0^2 = .01$	$\sigma_0^2 = .05$	1	1	0	0
	$\sigma_0^2 = .05$	$\sigma_0^2 = .10$	1	0	0	0
1981-82	MCOLS	$\sigma_0^2 = .01$	3	7	2	5
	$\sigma_0^2 = .01$	$\sigma_0^2 = .05$	7	6	0	1

	Change		**Rank Differences (No. of Cases)**			
Year	**From**	**To**	**2**	**3**	**4**	**5 or more**
	$\sigma_0^2 = .05$	$\sigma_0^2 = .10$	1	0	0	0
1988-89	MCOLS	$\sigma_0^2 = .01$	5	3	1	6
	$\sigma_0^2 = .01$	$\sigma_0^2 = .05$	4	3	0	0
	$\sigma_0^2 = .05$	$\sigma_0^2 = .10$	3	0	0	0

A1. List of Traditional Industries

Sl.No	ISIC	Name of Industries
1.	206	Mfg. and Refining of Sugar. C
2.	207	Production of Indigenous Sugar, Gur, etc. B
3.	211	Mfg. of Edible Oil, etc. A
4.	212	Tea Processing. A
5.	230	Cotton Ginning, Cleaning & Baling. B
6.	231	Cotton Spinning, Weaving, etc. A
7.	232	Printing Dyeing & Bleaching of Cotton Textile. A
8.	234	Production of Khadi. C
9.	235.	Weaving & Finishing of Cotton Textile in Handlooms. D
10.	236	Weaving & Finishing of Cotton Textile in Powerlooms.
11.	239.	Cotton Textiles Not Elsewhere Classified. C
12.	240.	Wool Cleaning,Baling & Pressing. A
13.	241.	Wool Spinning, Weaving & Finishing in Mills. A
14.	242.	Wool Spinning, Weaving & Finishing in other than Mills. C
15.	243	Dyeing & Bleaching of Woollen Textiles. B
16.	245	Spinning Weaving and Finishing of Silk Textiles. A

17.	246	Printing Dyeing and Bleaching of Silk Textiles. D
18.	251	Jute & Mesta Spinning and Weaving. D
19.	259	Mfg. of Jute Bags & Other Jute Textiles Not Elsewhere Classified.
20.	290	Tanning, Curing, Finishing, etc. of Leather. B
21.	291	Mfg. of Footwears(Except Vulcanised/Moulded Rubber Foot Wears). D
22.	292	Mfg. of Wearing Apparel of Leather & Substitutes. B
23.	293	Mfg. of Leather Consumer Goods (Other Than Apparel Foot Wears). B
24.	299	Mfg. of Leather & Fur Products Not Elsewhere Classified. C
25.	306	Products of Coal Tar in Coke Ovens. D
26.	307	Mfg. of Other Coal & Coal Tar Products Not Elsewhere Classified. B
27.	330	Iron & Steel. C
28.	331	Foundries for casting and Forging of Iron & Steel. B
29.	383	Mfg. of Jewellery and Related Articles. A

A2. List of Modern Industries

Sl.No.	ISIC	Name of Industries
1.	310	Mfg. of Basic and Industrial Organic & Inorganic Chemicals & Gases. B
2.	311	Mfg. of Fertilizer & Pesticides. C
3.	312	Mfg. of Paints, Varnishes and Lacquers. D
4.	313	Mfg. of Drugs and Medicines. A
5.	314	Mfg. of Perfumes, Cosmetics, etc. D
6.	315	Mfg. of Inedible Oils. B
7.	316	Mfg. of Turpentine, Synthetic Resins. A
8.	318	Mfg. of Explosives, Ammunitions & Fire Works. B
9.	319	Mfg. of Chemical Products Not Elsewhere Classified. C
10.	350	Mfg. of Agricultural Machineries & Equipments & Parts.
11.	351	Mfg. & Repair of Drills, Coal-Cutting Machines, etc. B
12.	352	Mfg. of Prime Movers, Boilers, etc. D
13.	353	Industrial Machineries for Food & Textiles. B
14.	354	Industrial Machineries Other Than for Food & Textiles.
15.	355	Mfg. of Refrigerators, Airconditioners, etc. A
16.	356	Mfg., Alteration & Repairs of Non Electrical Machineries, etc.
17.	357	Mfg. of Machine Tools. B
18.	359	Mfg & Repair of Non Electrical Machineries, etc. Not Elsewhere Classified.
19.	360	Mfg. of Electrical Industrial Machineries & Apparatus & Parts. A
20.	361	Mfg. of Insulated Wires & Cables. C
21.	362	Mfg. of Dry & Wet Batteries. A

22. 363 Mfg. of Electrical Apparatus,Appliances & Other Parts. C
23. 364 Mfg. of Radio, Televisions, etc. A
24. 366 Mfg. of Electronic Computers, Control Instruments.
25. 367 Mfg. of Electronic Components & Accessories Not Elsewhere Classified.
26. 369 Mfg. of Electrical Machineries, Apparatus & Appliances Not Elsewhere Classified. A
27. 374 Mfg. of Motor Vehicles & Parts. B
28. 375 Mfg. of Motorcycles, Scooters & Parts. B
29. 380. Mfg. of Medical, Surgical & Scientific Instruments. C

References

Afriat, S.N. (1972): 'Efficiency Estimation of Production Functions', *International Economic Review, 13, 568-598.*

Aigner, D.J. and S.F. Chu (1968): 'On Estimating the Industry Production Function', *American Economic Review, 58, 826-839.*

Aigner, D.J., C.A.K. Lovell and P. Schmidt (1977): 'Formulation and Estimation of Stochastic Frontier Production Function Models', *Journal of Econometrics, 6, 21-37.*

Battese, G.E. and T.J. Coelli (1988): 'Prediction of Firm Level Technical Efficiency with a Generalised Frontier Production Function And Panel Data', *Journal of Econometrics, 38, 387-399.*

Battese, G.E. and T.J. Coelli (1991): 'Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India', *Working Paper in Econometrics and Applied Statistics, No. 56, Dept. of Econometrics, University of New England.*

Bauer, P.W. (1990): 'Recent Developments in the Econometric Estimation of Frontiers', *Journal of Econometrics, 6,39-56.*

Broeck, J.Van Den, F.R. Forsund, L. Hjalmarsson and W. Meeusen (1980): 'On the Estimation of Deterministic and Stochastic Frontier Production Functions: A Comparison', *Journal of Econometrics, 13, 117-138.*

Cornwell, C., P. Schmidt and R.C. Sickles (1990): 'Production Frontiers With Cross Sectional and Time Series Variation in Efficiency Levels', *Journal of Econometrics, 6, 185-200.*

Debreu, G. (1951): 'The Coefficient of Resource Utilization', *Econometrica, 19, 272-292.*

Farrell, M.J. (1957): 'The Measurement of Productive Efficiency', *Journal of Royal Statistical Society A, 120, 253-290.*

Forsund, F.R. and L. Hjalmarsson (1979): 'Frontier Production Functions and Technical Progress: A Study of General Milk Processing in Swedish Dairy Plants', *Econometrica, 47,883-900.*

Forsund, F.R. and E.S. Jansen (1977): 'On Estimating Average and Best Practice Homothetic Production Functions via Cost Functions', *International Economic Review, 18, 463-476.*

Forsund, F.R., C.A.K. Lovell, and P. Schmidt (1980): 'A Survey of Frontier Production Functions and Of Their Relationship to Efficiency Measurement', *Journal of Econometrics, 13,5-25.*

Fried, H. O., C. A. K. Lovell and S. S. Schmidt (1993): 'The Measurement of Productive Efficiency: Techniques and Applications', *Oxford University Press, New York.*

Gabrielson, A. (1975): 'On Estimating Efficient Production Functions', *Working Paper No. A-35, Chr. Michelsen Institute, Department of Humanities & Social Sciences, Bergen, Norway.*

Ghosh, B. and C. Neogi (1993): 'Productivity, Efficiency and New Technology: The Case of Indian Manufacturing Industry', *Forthcoming in: The Developing Economies, Tokyo, Japan.*

Greene,W.H.(1980): 'Maximum Likelihood Estimation of Econometric Frontier Function', *Journal of Econometrics, 13,27-56.*

Greene, W.H. (1990) : 'A Gamma Distributed Stochastic Frontier Model', *Journal of Econometrics, 46, 141-163.*

Jondrow, J., C.A.K. Lovell, I.S. Materov and P. Schmidt (1982) : 'On the Estimation of Technical Efficiency in the Stochastic Frontier Production Function Model', *3Journal of Econometrics1, 19, 2/3, 233-238.*

Kalirajan, K.P. & J.C. Flinn (1983) : 'The Measurement of Farm-Specific Technical Efficiency', *3Pakistan Journal of Applied Economics1, 2, 167-180.*

Kopp, R.J. (1981): 'The Measurement of Productive Efficiency: A Reconsideration', *Quarterly Journal of Economics, 96, 477-503.*

Kumbhakar, Subal C. (1990): 'Production Frontiers, Panel Data and Time-Varying Technical Inefficiency', *Journal of Econometrics, 46, 201-211.*

Kumbhakar, Subal C. (1991): 'Estimation of Technical Inefficiency in Panel Data Models With Firm and Time-Specific Effects', *Economics Letters, 36, 43-48.*

Lee, L.F. (1983) : 'A Test for Distributional Assumptions for the Stochastic Frontier Functions', *Journal of Econometrics, 22, 245-267.*

Lee, L.F. and W.G. Tyler (1978) : 'The Stochastic Frontier Production Functions and Average Efficiency: An Empirical Analysis', *Journal of Econometrics, 7, 385-389.*

Meeusen, W. and J. Van Den Broeck (1977) : 'Efficiency Estimation From Cobb-Douglas Production Functions With Composed Error', *International Economic Review, 18, 435-555.*

Nishimizu, M., and J.M. Page, Jr. (1982) : 'Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia, 1965-78', *The Economic Journal, 92, 920-936.*

Pal, M., C. Neogi and B. Ghosh (1992) : 'On the Solution of Minimum Absolute Deviations Frontier Production Functions', *ISI Technical Report No. ERU/2/92.*

Pal, M., C. Neogi and B. Ghosh (1996) : 'Estimation of Frontier Production Function With Errors-in-Variables: An Illustration With Indian Industries', in *"Econometrics: Theory and Practice", Essays in honour of Professor N. Bhattacharya; Chakravarty et al. (eds.), Allied Publishers, New Delhi.*

Pettypool, M.D. and M.D. Trout (1988) : 'Recent Improvements to Data Envelopment Analysis', *Mathematical and Computer Modelling, 11, 1104-1106.*

Pitt, M. & L.F. Lee (1981) : 'The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry', *Journal of Development Economics, 9, 43-64.*

Richmond, J. (1974) : 'Estimating the Efficiency of Production', *International Economic Review, 15,515-521.*

Schmidt, P. (1976) : 'On the Statistical Estimation of Parametric Frontier Production Functions', *Review of Economics and Statistics, 58, 238-239.*

Schmidt, P. & T.F. Lin (1984) : 'Simple Tests of Alternative Specifications in Stochastic Frontier Models', *3Journal of Econometrics1, 24, 349-361.*

Schmidt, P. and R.C. Sickles (1984) : 'Production Frontiers and Panel Data', *Journal of Business and Economic Statistics, 4, 367-374.*

Seiford, L.H. and R.M. Thrall (1990) : 'Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis', *Journal of Econometrics, 46, 7-38.*

Sengupta, J.K. (1990) : 'Transformations in Stochastic DEA Models', *Journal of Econometrics, 46, 109-123.*

Stevenson, R.E. (1980) : 'Likelihood Functions for Generalised Stochastic Frontier Estimation', *Journal of Econometrics, 13, 57-66.*

Timmer, C.P. (1971) : 'Using a Probabilistic Frontier Function to Measure Technical Efficiency', *Journal of Political Economy, 79, 776-794.*

Arthaniti 2 (1-2), 62-78 / 2003

On Rationalization of Nash and Subgame-Perfect Behaviour from Observed Outcomes*

Indrajit Ray** and Susan Snyder***

Abstract

We discuss necessary and sufficient conditions, proposed by Sprumont (2000) and Ray and Zhou (2001), for observed outcomes in game forms, in which preferences are unobserved, to be rationalized as equilibria We use these conditions to analyze the rationalization by Nash and subgame-perfect equilibrium of data sets from a simple two player game in its normal and extensive form.

JEL Classification : C72, C92.

Keywords : Revealed Preference, Consistency, Nash Equilibrium, Subgame-Perfection

1. Introduction

Almost half a century back, social-choice theorists analyzed the concept of revealed preference *a la* Samuelson (1938) in the context of individual choices. The question asked was: given a series of observations of choices and choice sets, when can an individual's choice be rationalized by a preference ordering? It is now well-known that a choice function (possibly set-valued) can be rationalized by a complete and transitive preference relation if and only if it satisfies the so-called α and β conditions (Chernoff 1954, Arrow 1959, Richter 1966, Sen 1971). Many scholars have extended the revealed preference theory to more general settings (for example, Richter 1966); however up until recently, the literature has not considered the case of multiple individuals.

The question can naturally be asked in the context of interactive decisions. One could observe outcomes of game forms, in which preferences are unobserved, and then determine if each player's choices could be rationalized with a preference ordering, given some concept of equilibrium play. A number of recent papers considered this research agenda (Zhou 1997, Sprumont 2000, Ray and Zhou 2001, Sprumont 2001, Bossert and Sprumont 2002, Zhou 2002, Bossert and Sprumont 2003, Carvajal 2003, Carvajal, Ray and Snyder 2004, Ray and Snyder 2003).

The general problem under consideration in this literature takes the following form. Let Γ be a collection of related game forms that several individuals play. Suppose that we can observe an (or possibly, a set of) outcome(s), $O(G)$ for every game form G in Γ. What conditions must the observed outcomes satisfy so that $O(G)$ coincides with the equilibrium outcome (set) of a game associated with G, for every G? No restrictions are imposed on preferences except that they must be orderings. From the observations, one has to construct preferences of individuals so that the observed outcomes can be rationalized by the equilibrium notion with these constructed preferences.

---

* This is a part of an on-going project which was started when both authors were at Brown University. This paper borrows from other work done by the authors and their co-authors.

** Author for Correspondences Department of Economics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK E-mail: i.ray@bham ac uk; Fax. +44.121.414.7377.

*** Department of Economics, Brown University, Providence, RI 02912, USA. E-mail: Susan_Snyder@brown.edu; Fax: +1.401.863.1970.

Players can move simultaneously or sequentially in strategic situations. Thus, one could investigate two main types of game forms: normal games forms and extensive game forms. There are a variety of theoretical equilibrium concepts one can possibly analyze in each of these two game forms. Clearly, different structures of game forms and different solution concepts will have different implications for the observed outcomes.

We will concentrate mainly on two major solution concepts: Nash equilibrium for normal game forms and subgame-perfect equilibnum for extensive game forms. Sprumont (2000) first considered the issue for normal game forms and provided necessary and sufficient conditions for a set of observed outcomes to be Nash rationalizable. As a complement to the work of Sprumont, Ray and Zhou (2001) considered extensive game forms in which the players move sequentially with perfect information and provided necessary and sufficient conditions for a set of observed outcomes to be rationalizable by subgame-perfect equilibrium.

In this paper, we discuss the necessary and sufficient conditions, proposed by Sprumont (2000) and Ray and Zhou (2001). We consider a simple two player game in its normal and extensive form. We analyze all possible sets of observed outcomes for this game form and show where exactly these conditions are violated.

2. Set Up

2.1. Nash Behavior in Normal Games (Sprumont, 2000)

Let us consider finite sets of actions, A_i, one for each player, i; the product set, A, is called the set of joint actions. Let $\tilde{A}$ denote the set of all non-empty subsets of A which are Cartesian products; a typical element of $\tilde{A}$ is written as $B = \Pi_i B_i$, where $\emptyset \neq B_i \subseteq A_i$. We will call each B a *reduced normal game form*. A joint choice function, f, assigns to every reduced normal game form B a non-empty set $f(B) \subseteq B$. The joint choice function f may be a set, i.e., may not be singleton-valued. A data set is a realization of a joint choice function. A data set is *Nash rationalizable* if there exist preference orderings on A such that for every B, $f(B)$ coincides with the set of Nash equilibria for the game defined by the set of actions B with those preferences.

A *line* is defined as a subset of A that takes the form $\{(a_i, a_{-i}| a_i \in A_i\}$, where i and a_{-i} are arbitrary but fixed. For any $B, B' \in \tilde{A}$, denote by $B \vee B'$, the smallest set in $\tilde{A}$ that contains $B \cap B'$.

Condition 1. *Persistence under Expansion (PE).* For all $B, B' \in \tilde{A}$, $f(B) \cap f(B') \subset f(B \vee B')$.

This condition is a necessary condition for f to be Nash-rationalizable; however, it is not sufficient. Sprumont provides the following condition which together with the above generates the desired result as the following theorem suggests.

Condition 2. *Persistence under Contraction (PC).* For all $B, B' \in \tilde{A}$, such that $B' \subset B$, $f(B) \cup B' \subset f(B')$. If, moreover, B is included in a line of A and $f(B) \cup B' \neq \varnothing$, then $f(B') \subset f(B)$.

Theorem 1. (Sprumont, 2000) An n-agent joint choice function is Nash-rationalizable if and only if it satisfies Persistence under Expansion and Persistence under Contraction.

2.2. Subgame-Perfect Behavior in Extensive Games (Ray and Zhou, 2001)

Fix an extensive game form (tree) G with complete information. An extensive game form G is a finite rooted tree with set of nodes, X, with a distinct initial node and a precedence function $p: X/x_0 \rightarrow X$. If $p(y) = x$, then x is called an immediate predecessor of y. Also y is called an immediate successor of x, or $y \in s(x)$. Let $S(x)$ denote the set of all successors of x. A node z is called a terminal node, or an outcome, if there exists no $x \in X$ such that $p(x) = z$. The set of all terminal nodes is Z. A path $\wp$ is a finite sequence of nodes: $(x_k: k = 0, \ldots, m)$ where $x_k = p(x_{k+1})$ for each k and x_m is a terminal node. A path leading to a terminal node x_m, $\wp(x_m)$, can be uniquely identified.

A *reduced extensive game form* G' of an extensive game form G is an extensive game form consisting of (i) terminal nodes $Z' \subseteq Z$ and (ii) all the non-terminal nodes that belong to $\wp(z')$ for any $z' \in Z'$. For each reduced extensive game form G' and a non-terminal node $x \in X'/Z'$, the subgame form beginning at x, G_x', is the reduced extensive game form consisting of (i) terminal nodes $Z'(x) = Z' \cup S(x)$ and (ii) all the non-terminal nodes that belong to $\wp(z')$, for any $z' \in Z' \cup S(x)$.[1] The set of all such reduced game forms is Γ.

For Ray and Zhou, the data are the outcomes of all possible reduced game forms: formally, a data set is a realization of an outcome function $O: \Gamma \rightarrow Z$. They assume only one outcome is observed for each game. A data set is *subgame-perfect rationalizable* if there exist strict preference orderings on Z such that for every G',

---

[1] The subgame form G_x' is the reduced game form consisting of the path from x_0 to x and the subgame below the node x.

$O(G')$ coincides with the subgame-perfect equilibrium for the game defined by G' with those preferences.

Given an outcome function O, one can construct incomplete preference orderings for players over the terminal nodes. Consider the paths that lead to two different terminal nodes u and v. Take the player i who has to play at the node where these two paths diverge. Player i's preference over u and v can be determined by his choice in the reduced game form G' which has only two terminal nodes, u and v. This incomplete order, P_i, for player i, is known as the revealed base relation. Formally, for any $u, v \in Z$, let x be the node at which the paths to u and v diverge. If $x \in X_i$, then uP_iv if and only if $u = O(G')$, where G' is the reduced game form which has only two terminal nodes, u and v.

Ray and Zhou (2001) proposed three independent conditions that together turn out to be necessary and sufficient for subgame-perfect rationalization.

Condition 3. *Acyclicity of the Revealed Base Relation (AC).* For each player i, the revealed base relation, P_i, is acyclic.

Under Condition 3, P_i actually can be extended, via Zorn's lemma, to a strict preference ordering on Z which is both complete and transitive.

Condition 4. *Internal Consistency (IC).* For any G', if x is a predecessor of $O(G')$, then $O(G_x') = O(G')$.

Condition 5. *Subgame Consistency (SC).* Consider any G', $x \in X/Z$ and $u = O(G_x')$. For any $y \in s(x)$ such that y is not on the path to u and $v = O(G_y')$, then $O(G') = u$, where G' is the reduced-game form which has only two terminal nodes, u and v.

Theorem 2. (Ray and Zhou, 2001) An outcome function is subgame-perfect rationalizable if and only if acyclicity of the revealed base relation, internal consistency, and subgame consistency are all satisfied.

3. Illustration

In this section, we are going to analyze a simple two player game form. Consider the extensive game form in Figure 1a below. The main tree has two choice nodes; player 1 moves at the first node and has two choices, L and R. Player 2 moves at the

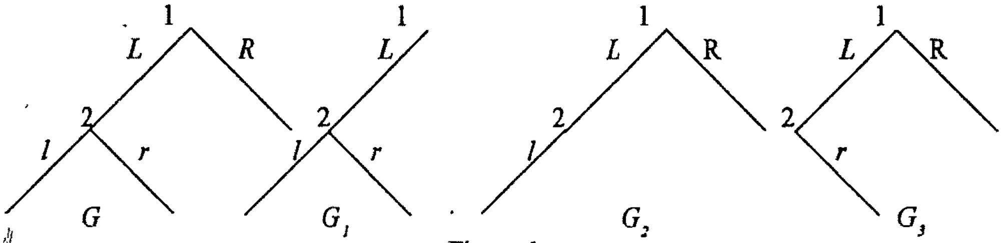


Figure 1a

second node (after player 1 moves L) and also has two choices, l and r. Here, the game form G has 3 possible (non-trivial) reduced game forms, denoted respectively by G_1, G_2 and G_3 in Figure 1a.

Consider now the corresponding normal game form, A. It has a 2x2 structure as shown in Figure 1b. There are however, four possible (non-trivial) reduced normal game forms; three of them correspond to the reduced extensive games forms G_1, G_2, and G_3: B_1, B_3 and B_4. Clearly, if we observe the extensive game outcomes, we do not observe an outcome corresponding to B_2, player 2's choice of action when player 1 chooses to play R. Note also that there are four possible strategy outcomes of the normal game form A, while there are only three possible outcomes of the extensive game form G.

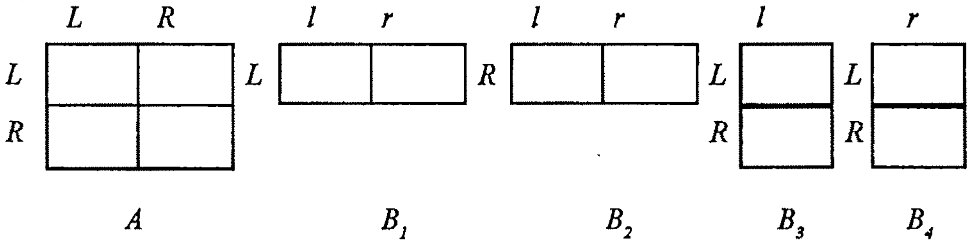


Figure 1b

This game form demonstrates an obvious distinction in the nature of the data sets observed in the simultaneous and the sequential game forms. In extensive game forms, we assume that we observe outcomes and not strategies (complete plans of actions), whereas in normal game forms, strategies (equivalently, actions) are assumed to be observed. Thus, a data set in the extensive form context has missing observations compared to the corresponding normal form data set. A normal form data set, if completely observed, is a richer data set.

We can list all the different data sets possible for game G and its reduced extensive game forms. For game G there are three possible outcomes, and for games G_1, G_2, and G_3 there are two possible outcomes each. Thus there are twenty-four possible data sets. These twenty-four cases are shown in Figures 1.1 – 1.24, where the bold lines indicate the outcome path.

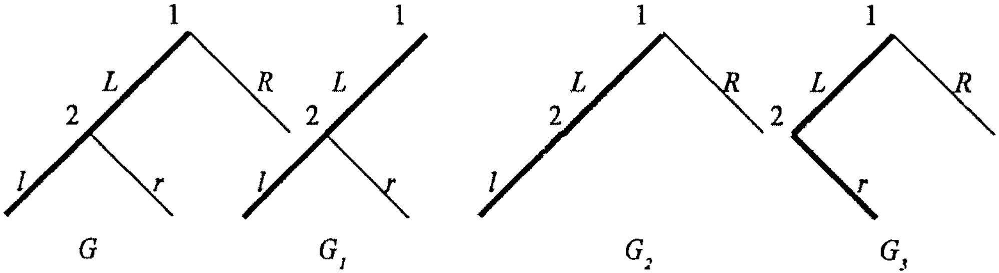


Figure 1.1

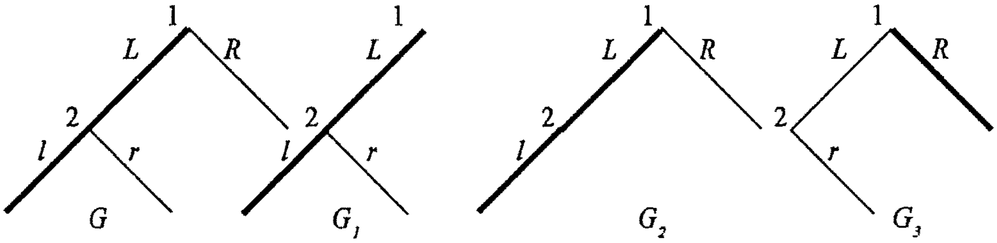


Figure 1.2

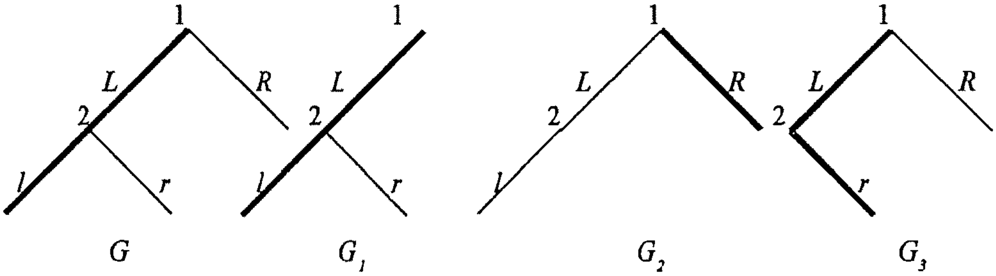


Figure 1.3

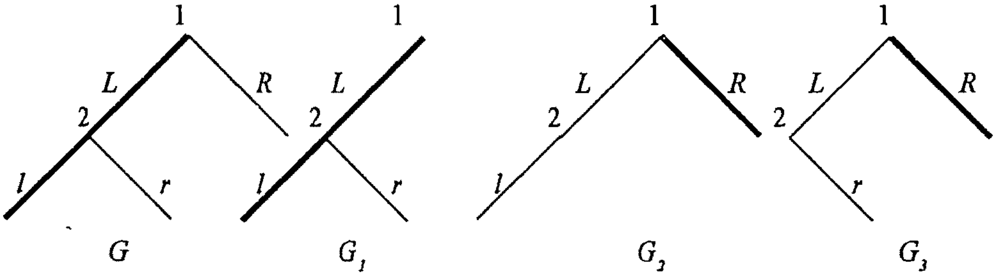


Figure 1.4

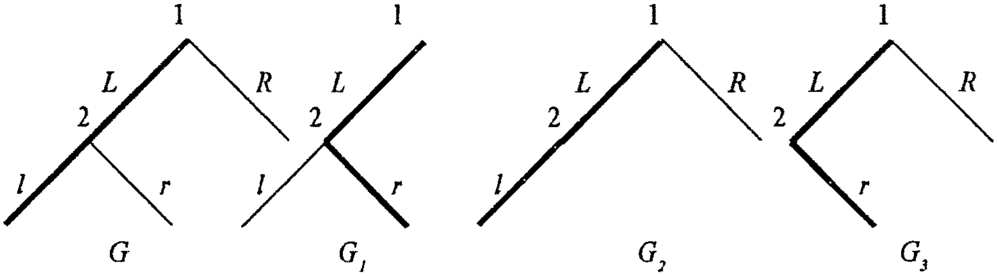


Figure 1.5

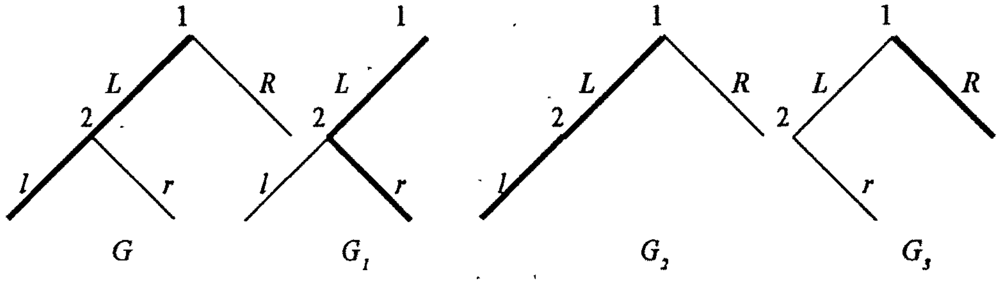


Figure 1.6

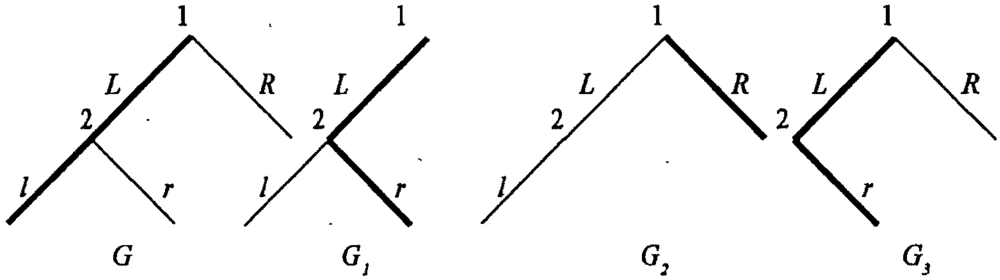


Figure 1.7

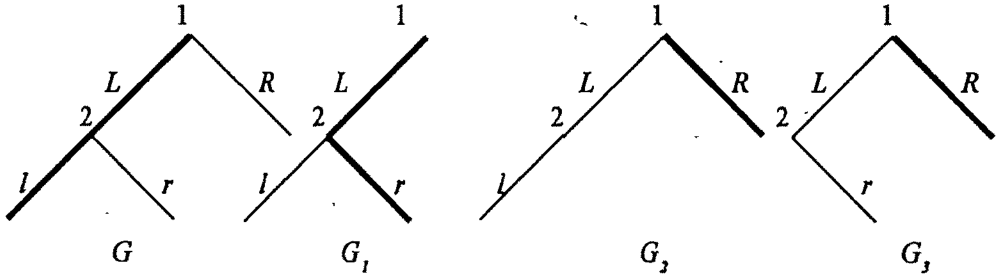


Figure 1.8

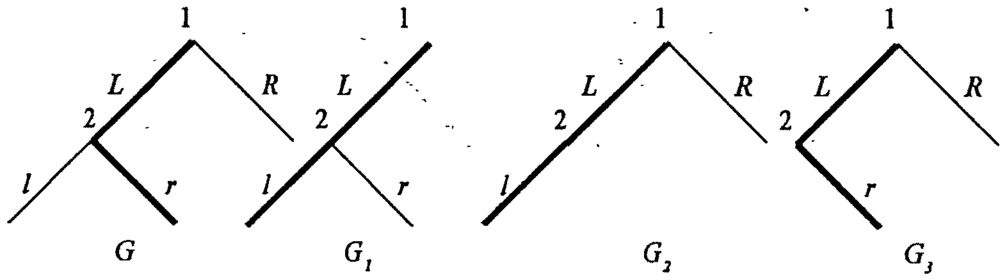


Figure 1.9

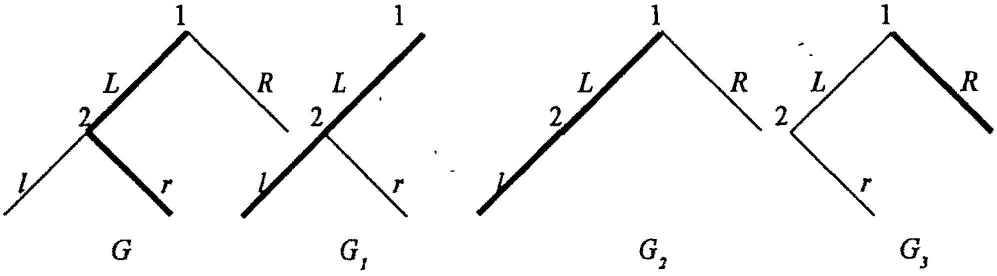


Figure 1.10

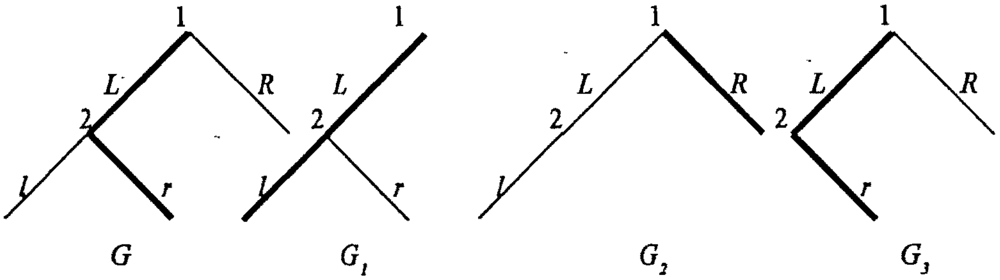


Figure 1.11

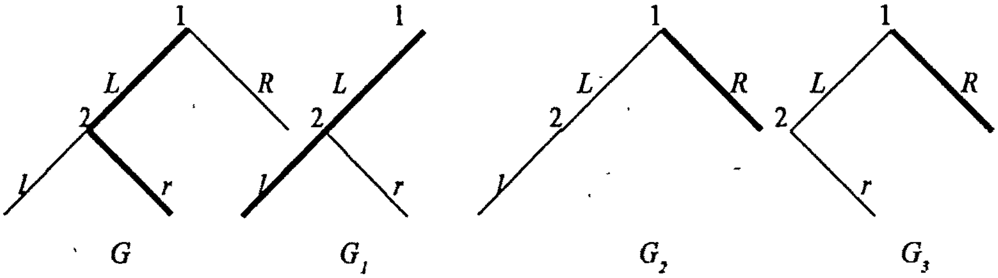


Figure 1.12

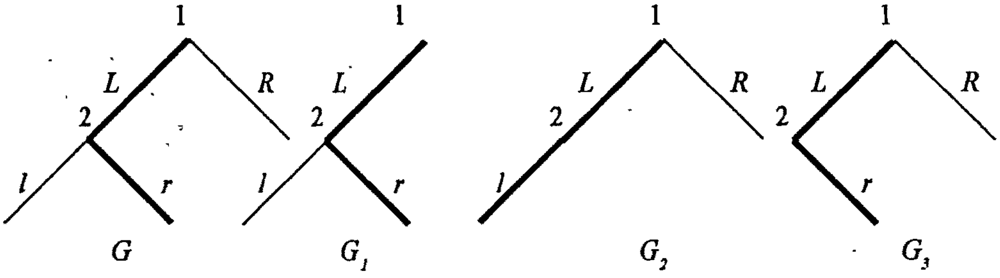


Figure 1.13

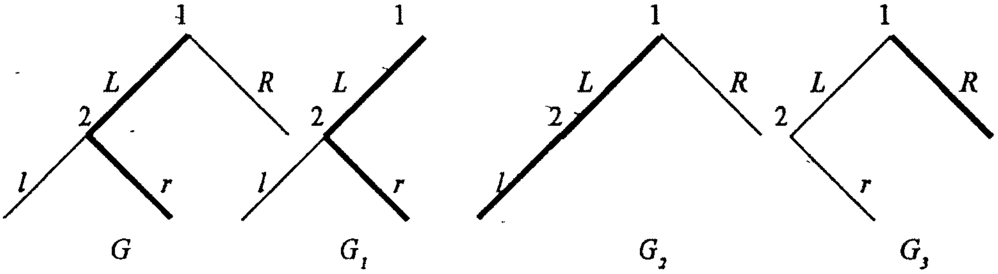


Figure 1.14

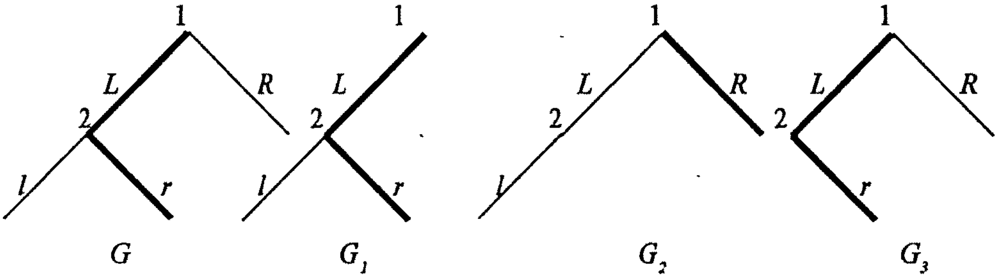


Figure 1.15

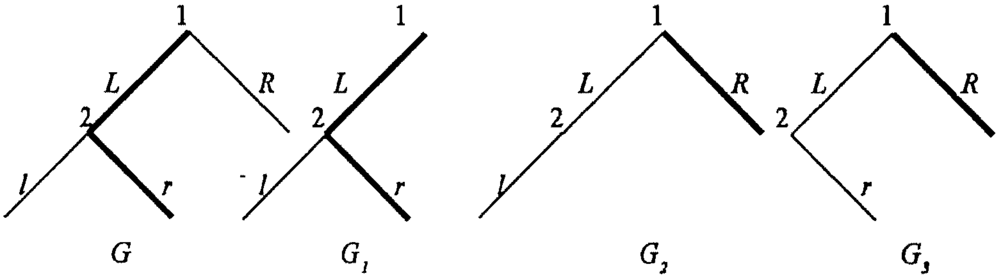


Figure 1.16

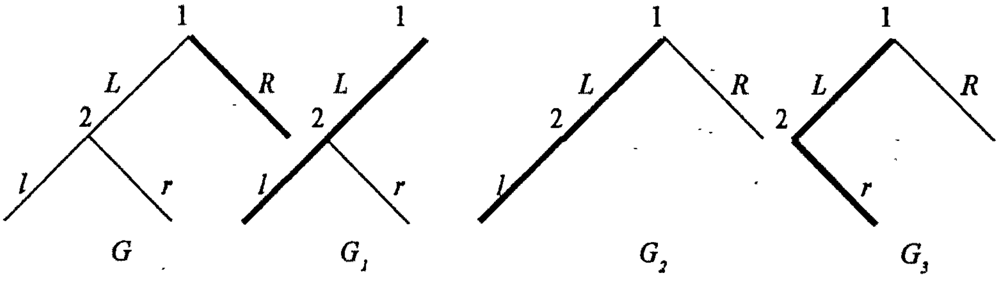


Figure 1.17

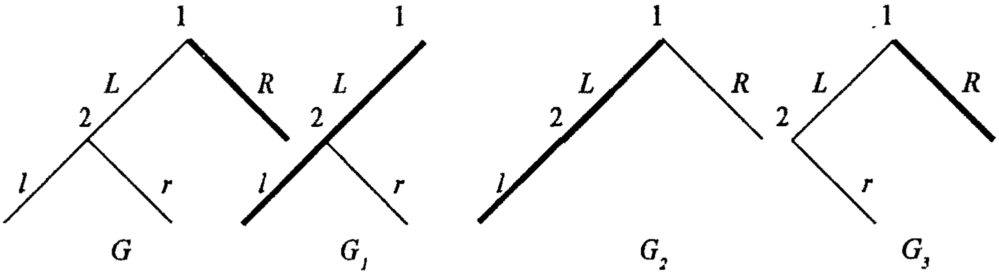


Figure 1.18

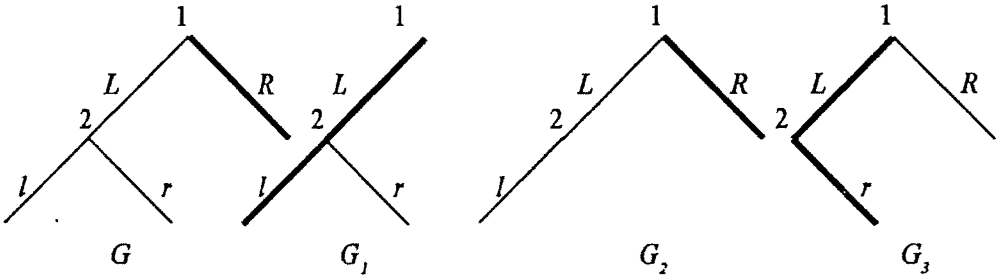


Figure 1.19

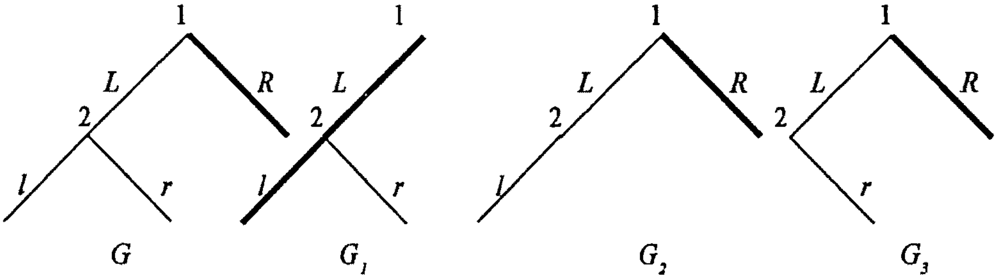


Figure 1.20

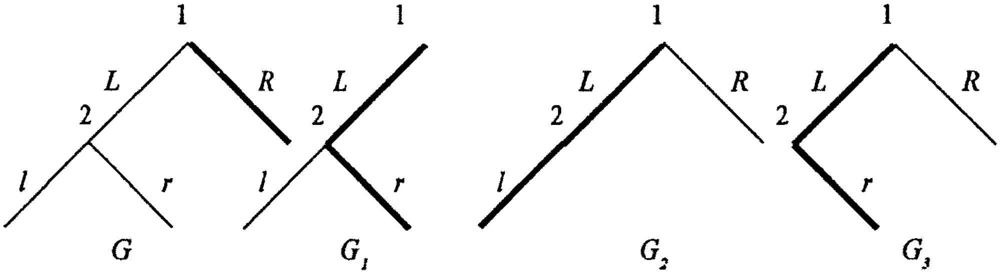


Figure 1.21

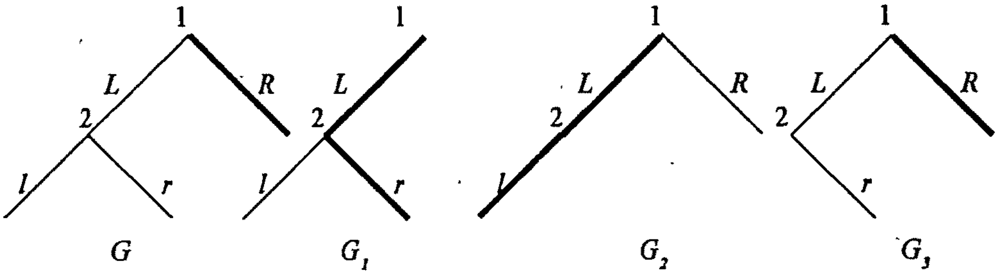


Figure 1.22

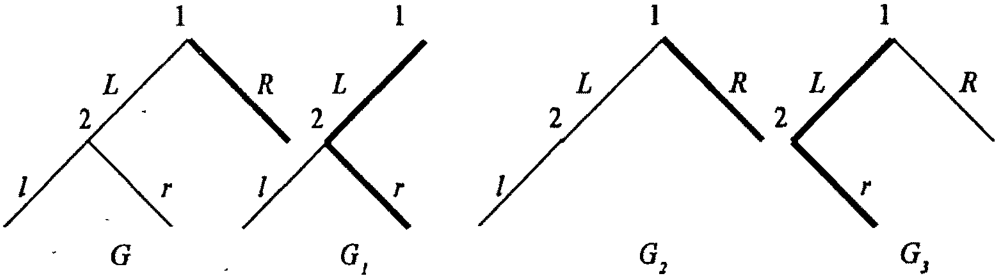


Figure 1.23

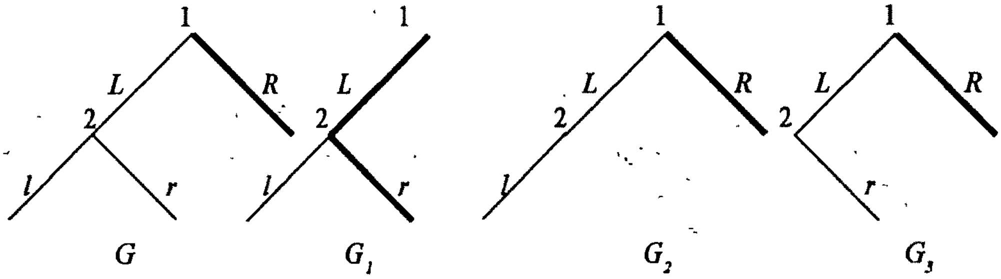


Figure 1.24

	Extensive Form Datasets		***Normal Form Datasets***			
Dataset	Consistent with subgame perfect?	Conditions violated	All data sets Consistent with Nash?	There exists data set consistent with Nash?	Conditions violated by all data sets	Conditions violated by some but not all data sets
1.1	YES		YES	YES		
1.2	YES			YES		PE
1.3		SC			PC	PE
1.4		SC			PC, PE	
1.5		IC			PC, PE	
1.6		IC			PC	PE
1.7		IC, SC			PC, PE	

	Extensive Form Datasets		*Normal Form Datasets*			
Dataset	Consistent with subgame perfect?	Conditions violated	All data sets Consistent with Nash?	There exists data set consistent with Nash?	Conditions violated by all data sets	Conditions violated by some but not all data sets
1.8		IC, SC			PC, PE	
1.9		IC			PC, PE	
1.10		IC, SC			PC, PE	
1.11		IC			PC	PE
1.12		IC, SC			PC, PE	
1.13	YES		YES	YES		
1.14		SC			PC	PE
1.15	YES			YES		PE
1.16		SC			PC, PE	
1.17		SC			PC, PE	
1.18		SC			PE	PC
1.19	YES			YES		PC, PE
1.20	YES			YES		PC, PE
1.21		SC			PC, PE	
1.22	YES			YES		PC, PE
1.23		SC			PE	PC
1.24	YES			YES		PC, PE

Table 1

In the second column of Table 1, we list whether play is consistent with subgame-perfect equilibrium for each data set from the extensive game forms, i.e., whether the outcome function in the extensive game form satisfies Ray and Zhou's conditions AC, IC, and SC. Eight of the twenty-four data sets are rationalizable as subgame-perfect play. In the next column, we list which conditions were violated, if any (note AC is never violated in this game form).

Suppose, instead, we observe data sets from the normal game form, A. To facilitate comparison with the extensive game form, we will assume that the outcome functions map to singletons, i.e., we will test for unique Nash equilibrium in each game. There are multiple normal form data sets consistent with a given extensive form data set. For example, suppose the observed outcome of game form G in the extensive game form data set is player 1 chooses L and player 2 chooses l. This corresponds to a unique strategy profile in the normal game form A: $f(A) = (L, l)$. Whether we observe this data from the extensive game form or the normal game form, the information is equivalent. However, if the outcome of the extensive form game G is that player 1 chooses R, then there are two possible strategy profiles in the normal form data set consistent with this observation: $f(A) = (R, l)$ and $f(A) = (R, r)$. Additionally, all normal form data sets will contain an observation of the choice from the reduced normal game form B_4, for which there is no corresponding extensive game form. In total, there are 2 multiple normal form data sets consistent with a given extensive form data set for data sets in Figures 1.1 - 1.16, and there are four normal form data sets associated with each extensive form data set for games in Figures 1.17 - 1.24.

Consider the extensive game form data set in Figure 1.1. As described above for G, and similarly, for the reduced extensive game forms G_1, G_2 and G_3, each outcome corresponds to a unique choice in each of the corresponding reduced normal game forms A, B_1, B_2 and B_3. Given this data, any outcome in B_4 is consistent with Nash equilibrium; i.e., both corresponding normal form data sets satisfy Sprumont's conditions of PC and PE. In the fourth column of Table 1, we indicate whether all corresponding normal form data sets are consistent with unique Nash equilibrium outcomes. There are only two cases where this occurs.

Generally the extra data available from the normal form may change the test results. The normal form data set in Figure 1.2 will be consistent with a unique Nash equilibrium in each game if $f(B_4) = (R, l)$, but not if $f(B_4) = (R, r)$, because PE is not satisfied in the second case. In the fifth column of Table 1, we indicate whether there exists any corresponding normal form data set consistent with unique Nash equilibrium.

The last two columns of the table indicate which, if any, of Sprumont's conditions are violated by (column six) all and (column seven) at least one but not all of the corresponding normal form data sets.

Note that tests using the extensive form data sets are strictly weaker: whenever the extensive form conditions are violated, the normal form conditions are violated in all cases. If at least one of the corresponding normal form data sets of an extensive form data set is rationalizable by unique Nash equilibria, then the extensive form data set is consistent with subgame-perfect equilibrium. This is not surprising as the normal form data sets contain extra information. On the other hand, the extensive form tests have a subgame-perfect restriction that the normal form tests do not.

We draw no other conclusions about general relationships between the extensive form conditions and the normal form conditions. It does seem clear that the relationship changes depending on whether the subgame-perfection condition is binding, that is, whether the extensive form outcomes can be Nash but not subgame-perfect. Subgame-perfection only restricts outcomes when player 1 moves R in game G, which occurs in data sets in Figures 1.17 - 1.24. Thus we see, for example, that in data sets in Figures

1.1 - 1.16, a violation of condition SC in the extensive form data set implies a violation of PC in all corresponding normal form data sets, but for data sets in Figures 1.17 - 1.24 a violation of condition SC in the extensive form data set implies a violation of PE in all corresponding normal form data sets.

4. CONCLUSION

4.1. Related Work

In the framework of normal game forms, Sprumont (2000) also discusses Pareto efficiency as a solution concept. The falsifiable implications of the Nash and Pareto hypotheses are potentially different. Sprumont considers two-agent joint choice functions which select a single joint action from each pair of feasible sets and proves that if any such function is Nash-rationalizable, it is also Pareto-rationalizable. Also, every choice function is partially Pareto-rationalizable, even if strict preferences are required.

Bossert and Sprumont (2003) add to this literature by exploring testable restrictions for (Pareto) efficiency and individual rationality. To understand individual rationality, they use the status quo model as introduced by Zhou (1997) and used by Rubinstein and Zhou (1997), and Mastalioglu and Ok (2003). They find a condition, namely, END-congruence which is necessary for arbitrary domains. This property is also sufficient in two cases: (i) when feasible sets under consideration coincide with the universal set and (ii) for arbitrary domains, when the cardinality of the universal set is smaller than $2(n + 1)$.

In this research agenda, it is generally required that the outcomes of all possible reduced trees be observable. There are many simple and interesting game situations, such as dividing-a-dollar, in which the players have infinitely many strategies but practically only finitely many observations can be made. Hence, one would like to know how to extend the results in this literature to such cases with only limited observations.[2] Carvajal (2003) attempts to model such a situation and finds testable conditions.

Ray and Snyder (2003) provide separate testable restrictions for Nash and subgame-perfect equilibrium. They identify two distinct conditions, such that the first one (Extensive Form Consistency) is necessary and sufficient for rationalization by Nash equilibrium behaviour and together with Subgame Perfect Consistency, is necessary and sufficient for rationalization by subgame-perfect equilibrium. The conditions together clearly are equivalent to the three conditions proposed by Ray and Zhou for subgame-perfect rationalization. The advantage however is that their conditions can be used to test for Nash behaviour alone and also to distinguish between

---

[2] This is parallel to Afriat's (1967) approach to the theory of revealed preferences, vis-à-vis Samuelson's.

Nash and subgame-perfect behaviour.

4.2. Future Research

It is important to establish and understand the mutual relationships among all the conditions proposed by Sprumont (2000), Ray and Zhou (2001) and Ray and Snyder (2003). The features of the different data sets we described in Section 3 of this paper suggests that the tests in the extensive game forms are some sort of projection from those in the corresponding normal game forms and therefore it might be useful to formulate all these restrictions in a semi-algebraic framework and see if one can uncover a relationship that way.

In this paper, we have studied two variants of equilibrium only. It would be interesting to extend this study to other solution concepts as well. If the observed outcomes cannot be rationalized by Nash equilibrium, can they be rationalized by other solution concepts, such as the maxmin solution? For 2×2 normal game forms, it is easy to derive a necessary and sufficient condition for maxmin rationalization.[3] Suppose, for each player, if his response to either of his opponent's strategies is the same (as revealed by the outcome of the problem in which his opponent has only one strategy available), then, intuitively, he should choose the same strategy even when his opponent has both of her strategies available. Let us call this condition *non-reversal of domination*. One can verify that, for 2×2 normal game forms, the observed outcomes can be rationalized by the maxmin solution if and only if the outcome function satisfies non-reversal of domination. For more general normal forms, this condition continues to be necessary, but no longer sufficient, as there are (rather tedious) counterexamples. More work is needed to resolve this issue.

Ray and Snyder's (2003) conditions, to distinguish Nash and subgame-perfect behaviour, are constructed in such a way that violations of their conditions refer specifically to players and nodes. Checking these conditions can help identify the players and the nodes where subgame-perfect or Nash behaviour is not observed. Thus, one may be able to recover information about individual (ir-)rationality. This could be relevant to obtain results to rationalize observed outcomes using other notions of rationality such as multiple rationales (Kalai, Rubinstein and Spiegler, 2002). It should look for general results analogous to Kalai, Rubinstein and Spiegler (2002) in this context.

This literature is clearly related to the studies in experimental game theory as both try to verify whether players in "real" situations play according to various equilibrium concepts. The important distinction, however, is that we assume that players' preferences are unknown while most of the experimental studies assume that players' utility functions are of specific forms. One could set up an experiment to see whether players do satisfy these conditions or not.

---

[3] We would like to thank Lin Zhou for suggesting this.

An important implication of this study would be for individual decision making. One could study dynamic inconsistency of naive and sophisticated agents (O'Donoghue and Rabin, 1999). Recently, Ok (2002) studied how one can represent possibly incomplete preference relations by a vector-valued utility function. One may think of incomplete preferences in our set up as well. One might want to rationalize observed outcomes using incomplete preferences.

References

Afriat, S. (1967): 'The Construction of a Utility Function from Demand Data', *International Economic Review 8, 67-77.*

Arrow, K. (1959): 'Rational Choice Functions and Orderings', *Economica 26, 121-127.*

Bossert, W., and Sprumont, Y. (2002): 'Core Rationalizability in Two-Agent Exchange Economies', *Economic Theory 20, 777-791.*

Bossert, W., and Sprumont, Y. (2003): 'Efficient and Non-Deteriorating Choice', *Mathematical Social Sciences 45, 131-142.*

Carvajal, A. (2003): 'On Individually-Rational Choice and Equilibrium', *PhD Thesis, Brown University.*

Carvajal, A., Ray, I., and Snyder, S. (2004): 'Equilibrium Behavior in Markets and Games: Testing, Identification, and Comparative Statics', *Journal of Mathematical Economics, Forthcoming.*

Chernoff, H. (1954): 'Rational Selections of Decision Functions', *Econometrica 22, 422-443.*

Kalai, G., Rubinstein, A., and Spiegler, R. (2002): 'Rationalizing Choice Functions by Multiple Rationales', *Econometrica 70, 2481-2488.*

Masatlioglu, Y., and Ok, E. (2003): 'Rational Choice with Status Quo Bias', *Mimeo, New York University.*

O'Donoghue, T., and Rabin, M. (1999): 'Doing it now or Later', *American Economic Review 89, 103-124.*

Ok, E. (2002): 'Utility Representation of an Incomplete Preference Relation', *Journal of Economic Theory 104, 429-449.*

Ray, I., and Snyder, S. (2003): 'Observable Implications of Nash and Subgame Perfect Behaviour in Extensive Games', *Working Paper 03-02, Department of Economics, Brown University.*

Ray, I., and Zhou, L. (2001) : 'Game Theory via Revealed Preferences', *Games and Economic Behavior 37, 415-424.*

Richter, M. (1966) : 'Revealed Preference Theory', *Econometrica 34, 635-645.*

Rubinstein, A., and Zhou, L. (1997): 'Choice Problems with a Reference Point', *Mathematical Social Sciences 37, 205-209.*

Samuelson, P. (1938) : 'The Empirical Implications of Utility Analysis', *Economica 6, 344-356.*

Sen, A. K. (1971) : 'Choice Functions and Revealed Preference', *Review of Economic Studies 38, 307-317.*

Sprumont, Y. (2000) : 'On the Testable Implications of Collective Choice Theories', *Journal of Economic Theory 93, 205-232.*

Sprumont, Y. (2001) : 'Paretian Quasi-Orders: The Regular Two-Agent Case', *Journal of Economic Theory 101, 437-456.*

Zhou, L. (1997) : 'Revealed Preferences and Status Quo Effect', *Mimeo, Duke University.*

Zhou, L. (2002) : 'Testable Implications of Nash Equilibrium Theory', *Mimeo, Arizona State University.*

Arthaniti 2 (1-2), 79-96 / 2003

Foreign market entry and host-country welfare : A theoretical analysis*

Arijit Mukherjee**

Abstract
Many developing countries are liberalizing their economies to allow higher equity participation by foreign firms. We argue that the possibility of joint venture may reduce the number of technology transfers Hence, joint venture may reduce the welfare of a host-country by creating higher market concentration. However, higher profit generation under joint venture encourages the foreign firm to transfer relatively better technology to the host-country and may make the host-country and the firms better-off under joint venture compared to licensing. We show that the host-country allows fully owned subsidiary of the foreign firm for sufficiently large efficiency-gain.

JEL Classification : F21, F23.

Keywords : Joint venture, Licensing, Welfare.

1. Introduction

Technical progress is an important factor for economic growth. Many developing countries are liberalizing their policies to attract superior technologies from the technologically advanced foreign firms. In the last few decades the cases of international technology transfers have increased significantly. However, initially the developing countries were mainly interested for technology licensing. For example, in the early 1980s India has opened the doors for seeking new technologies through licensing or technological collaborations and the number of such collaborations has grown (see e.g., Economic and Political Weekly, 1985 and Bagchi, 1987). However, equity participation by the foreign firms was continued to be restricted.

As time passes, these countries gradually relax their restrictions and allow foreign equity participation in the domestic projects. However, the policy reforms were targeted mainly for international joint ventures (JVs) than fully owned foreign subsidiaries. As an example, India, in the early 1990s, has allowed 51% foreign equity participation in 35 high priority sectors as automatic approval but for higher than this, it is decided through case-by-case study (see Saqib, 1995). The literature on international business also shows rising trend of international JVs in the developing countries. Korbin (1988)

---

* An informal discussion with Sudipto Bhattacharya on related issues has helped me to develop this idea. I would like to thank an anonymous referee of this journal and Rod Falvey for helpful comments and suggestions. I also acknowledge the financial support from the Netherlands Technology Foundation (STW) The usual disclaimer applies.

** School of Economics, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
E-mail: arijit.mukherjee@nottingham.ac.uk
Fax: +44-115-951 4159

reported that in some less developed countries nearly 80% of total foreign capital entered through JVs.

The present paper starts off with this background and highlights a potential conflict between the interests of the firms and the host-country. We show that if the foreign technology supplier cannot commit to exclusive technology licensing, JV between the technology supplier and a technology buyer may eliminate the commitment problem and generate higher industry profit by increasing market-concentration. Therefore, while JV is preferable to the firms, welfare of the host-country may be higher under technology licensing. However, if the foreign firm transfers sufficiently superior technology to the host-country under JV compared to licensing, JV is profitable to both the firms and the host-country.

Therefore, the main message of this paper is following. If international JVs are formed mainly to get the benefits of higher market concentration, it may not be worthwhile to encourage such JVs. But, if higher market-concentration leads to sufficient efficiency-gain by attracting relatively better technology to the host-country, these JVs are beneficial to the host-country. We show that for sufficiently large efficiency-gain, the host-country is better-off if it allows fully owned subsidiary of the foreign firm.

The vast literature on international technology licensing and international JVs has addressed several issues such as the possibility of technology licensing, the quality of the transferred technology, relative benefits of licensing and foreign investment, impact of different government policies on the entry strategy of a foreign firm and impact of different entry strategies on host-country welfare. As a representative sample, one may look at Teece (1977, 1981), Svejnar and Smith (1984), Falvey and Fried (1986), Marjit (1990b), Tang and Yu (1990), Kabiraj and Marjit (1992, 1993), Lee and Shy (1992), Singh (1992), Purakayastha (1993), Al-Saadon and Das (1996), Das (1998), Marjit and Mukherjee (1998), Mukherjee (2000 and 2002) and Sinha (2001a and 2001b).[1] In the present paper we address a different problem associated with international technology transfer, viz., the inability of a foreign firm to commit to the exclusive technology transfer under licensing.

The present paper is closely related to Tang and Yu (1990) and Lee and Shy (1992). While examining different entry strategies of a foreign firm, Tang and Yu (1990) conclude that welfare of a host-country is maximal under international JV. Contrary to Tang and Yu (1992), the present paper shows that welfare of a host-country may be higher under licensing or fully owned subsidiary compared to JV if licensing involves opportunistic behavior of the foreign technology supplier and technology transfer is costly. Focusing on international JV only, Lee and Shy (1992) argue that whether an increase in the regulated share of foreign ownership increases welfare of the host-country depends on the managerial control in the JV. In contrast, the present paper focuses on the role of market concentration and associated efficiency-gain to determine the effect of government regulated foreign shareholding on the host-country welfare.

---

[1] Markusen (1995) and Pack and Saggi (1997) provide surveys on this area.

The remainder of the paper is organized as follows. Section 2 provides the basic model and the results. Section 3 concludes.

2. The model and the results

Consider a firm, called firm 1, who has technologies[2] to produce a product and wants to enter a country, called host-country. Firm 1 can enter the host-country either through technology licensing or foreign investment.[3] However, the government policy of the host-country imposes constraint on the shareholding of firm 1 if it wants to do foreign investment. Assume that there are two potential producers, called firm 2 and firm 3, in the host-country. Neither of the firms 2 and 3 has the technology to produce the product and assume that they are symmetric.

Consider the following game. At stage 1, the welfare[4] maximizing government of the host-country decides the maximum proportion of shareholding of firm 1 in the host-country. Define the shareholding of firm 1 by α . Hence, the host-country government decides that firm 1 can hold up to $\overline{\alpha}$ proportion of shareholding in the host-country, where $\overline{\alpha} \in [0,1]$. At stage 2, firm 1 decides about the quality of the technology to be transferred to the host-country. At stage 3, firm 1 decides the mode of technology transfer (i.e., either licensing or investing[5] in the host-country) and also the number of technology transfers.[6] At stage 4, the outputs are produced and the profits are realized. We solve the game through backward induction. The sequence of the moves of the game is shown in Figure 1.

So, we consider a game with four players: three firms and the host-country government. Host-country government's strategy space consists of the proportion of shareholding of firm 1. Firm 1's strategy space consists of the quality of the technology to be used in the host-country and also the mode of technology transfer, i.e., licensing or foreign investment. While deciding the mode of technology transfer, firm 1 gives a

---

[2] We measure the quality of the technology by the corresponding marginal cost of production. Better quality of the technology implies lower marginal cost of production.

[3] To explain the argument of this paper in the simplest way, we rule out the possibility of export by firm 1. Significant amount of transportation cost may make export unprofitable to firm 1.

[4] Welfare of the host-country is defined as the summation of the total profits of the host-country firm(s) and consumer surplus.

[5] It is trivial that if, in stage 1, the host-country government decides no shareholding of firm 1, licensing is the only option to firm 1.

[6] Note that there is an extensive form game involved while deciding on licensing or foreign investment, since the number of technology transfer is also a choice of firm 1. We define this extensive form game under licensing and foreign investment in subsections 2.1.1 and 2.1.2 respectively.

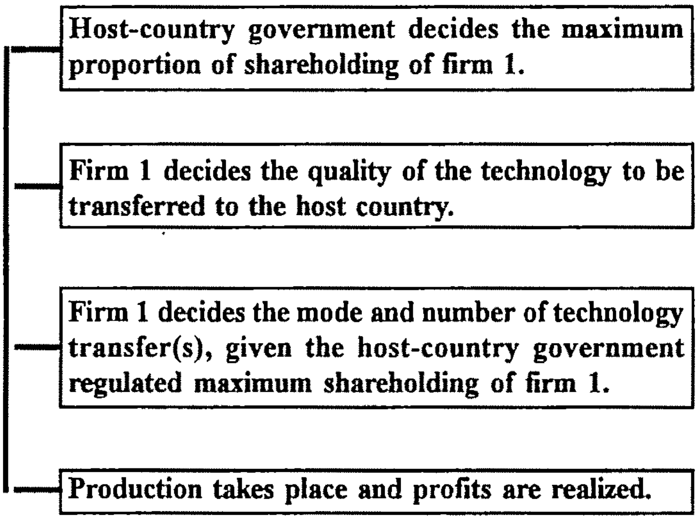

Host-country government decides the maximum proportion of shareholding of firm 1.
Firm 1 decides the quality of the technology to be transferred to the host country.
Firm 1 decides the mode and number of technology transfer(s), given the host-country government regulated maximum shareholding of firm 1.
Production takes place and profits are realized.

Figure 1 : The sequence of the moves of the game

take-it-or-leave-it offer to the host country firm(s), firms 2 and 3, except for the situation where firm 1 does foreign investment with a fully owned subsidiary. The strategy spaces of firms 2 and 3 involve either accepting or rejecting the offer of firm 1. Assume that each of firm 2 and firm 3 accepts the offer of firm 1 as long as acceptance does not make the concerned firm worse-off compared to rejection. We normalize the profits of the host-country firms to zero if they do not produce this product. Hence, each host-country firm accepts an offer of firm 1 if its net profit is positive.

We assume that there are no costs specific to licensing and foreign investment. This assumption helps us to look at the effects of market concentration on host-country policy. Also, assume that the host-country law does not allow firms 2 and 3 to collude.

Assume that the products of all the firms are homogeneous and the inverse market demand function for the product is

$$P = a - q, \tag{1}$$

where the notations have usual meanings. We assume that if there is more than one firm producing in the host-country, competition in the product market is characterized by Cournot competition.

Let us first examine the firm 1's incentive for licensing and foreign investment for given host-country policy and the quality of the technology. Then we will see the effects on the quality of the technology and the host-country policy.

2.1 For a given host-country policy and foreign technology

Suppose the host-country government decided a proportion $\overline{\alpha} \in (0,1)$ as the maximum shareholding of firm 1 and firm 1 has decided to transfer the technology corresponding to the constant marginal cost of production c. In the following analysis we define the profit of the ith firm in the product market (i.e., revenue minus total variable cost) by π_i, $i = 1,2,3$.

Note that if the host-country government allows fully owned subsidiary of the foreign firm, i.e., $\overline{\alpha} = 1$, firm 1 has no incentive for licensing or holding the proportion of share less than 1 in the host-country firm. Hence, in this situation, the foreign firm opens its fully owned subsidiary in the host-country. But, if the host-country government does not allow fully owned subsidiary of the foreign firm, the foreign firm has to decide between licensing and foreign investment with the maximum proportion of shareholding $\overline{\alpha} = 1$. We define this latter possibility as joint venture (JV).

2.1.1 Technology licensing

Let us consider the situation where firm 1 decides to license its technology to the host-country firm(s). We assume that *firm 1 cannot commit to an exclusive technology transfer*. This commitment problem implies that there is a possibility that firm 1 may license its technology to a host-country firm even if it licenses the technology to the other host-country firm. High cost of monitoring or lack of punishment possibility may encourage the firm 1 to transfer the technology to a host-country firm once it has transferred the technology to the other host-country firm.[7] Alternatively, we may think that the firm 1 can invent around the initial technology and make a similar technology, which corresponds to the constant marginal cost of production c.[8]

The extensive form game under licensing is as follows. First, firm 1 gives a take-it-or-leave-it licensing offer to one of the host-country firms. Without loss of generality, we assume that firm 1 gives the licensing offer to firm 2. Firm 2 either accepts or rejects the offer. If firm 2 accepts the offer of firm 1, firm 1 licenses the technology to firm 2 and takes an up-front fixed fee.[9] Then firm 1 decides whether to license the technology to firm 3 also. If firm 1 decides to give a licensing contract to firm 3, it gives a take-it-or-leave-it offer to firm 3 and firm 3 either accepts or rejects the offer. If firm 2 has previously rejected the offer of firm 1, firm 1 gives a take-it-or-leave-it licensing offer to firm 3. Upon acceptance or rejection by firm 3, firm 1 decides whether to give a take-it-or-leave-it licensing offer to firm 2, and the game continues in this way until firm 1 stops offering further licensing contract. Firm 2 or firm 3 accepts an offer as long as acceptance does not make the concerned firm worse-off compared to rejection.[10] Hence, we consider alternating take-it-or-leave-it licensing offers by firm 1 to firm 2 and firm 3 until firm 1 stops offering licensing contract. We assume away any cost of delaying between the offers.

---

[7] See Hart and Tirole (1990) for related problems in the vertically separated industry.

[8] The literature on licensing in the oligopolistic markets has considered the possibility of re-entry by the licenser after technology licensing. For example, one may look at Katz and Shapiro (1985), Marjit (1990a) and Mukherjee (2001).

[9] The problem of getting ex-post payment under licensing may induce the firm 1 to charge only up-front fixed fee under patent licensing contract. This may be due to purely informational reasons or because of the imitation possibility from the licensee after the technology transfer. For example, one may refer to Katz and Shapiro (1985), Marjit (1990a), Marin and Schnitzer (1995) and Mukherjee (2001) for licensing contract with fixed-fee only.

Proposition 1: *Firm 1 licenses the technology to firm 2 and firm 3 and firm 1's payoff is* $2\pi(c,c)$.[11]

Proof: It is easy to understand that, under the problem of licensing specified above, firm 1 has always the incentive to license the technology to another firm given that it has already licensed the technology to a firm. Thus, if both firm 2 and firm 3 have the licensed technology, then each licensee gets a payoff $\pi(c,c)$. Hence, firm 2 and firm 3 accept an offer which calls for an up-front fixed fee less than or equal to $\pi(c,c)$; otherwise, these firms are better-off by rejecting an offer which calls for an up-front fixed fee more than $\pi(c,c)$. Therefore, firm 1 charges $\pi(c,c)$ to each licensee and licenses the technology to firm 2 and firm 3. Q.E.D.

2.1.2 Foreign investment

Now, consider another way of technology transfer, viz., foreign investment. Since, it is clear that firm 1 will open its fully owned subsidiary when the host-government allows $\bar{\alpha}$ to be one, in this subsection we restrict our attention to those foreign investments where $0<\bar{\alpha}<1$, i.e., on JV between firm 1 and a host-country firm. In case of a JV, we assume that firm 1 transfers the technology to the JV firm and the partners receive the shares of profits according to their equity shares in the JV (see, e.g., Tang and Yu, 1990, Lee and Shy, 1992, Al-Saadon and Das, 1996).[12]

With the possibility of JV, we consider a game like the previous subsection with the exception that, at the beginning, firm 1 gives a take-it-or-leave-it JV offer to firm

---

[10] If one interprets the technology of the foreign firm as input then our framework has a similarity with the one considered in Hart and Tirole (1990) if the fully owned subsidiary of the foreign firm in the host-country is not allowed. However, the present paper focuses on a problem, viz., on the host-country policy, which is different from Hart and Tirole (1990), and endogenizes the possibility of licensing, partial integration and fully owned subsidiary of the foreign firm.

[11] Call $\pi_2(c,c)=\pi_3(c,c)=\pi(c,c)$, where the first and second arguments of the profit functions show the marginal cost of firm 2 and firm 3 respectively.

[12] Profits are assumed to be verifiable. Further, we assume away the possibility of other types of payments like up-front fixed fee and/or output royalty along with equity sharing in the JV. Often, the government regulation may be a reason for this (see Tang and Yu, 1990). However, in our analysis, the possibility of other type of payment such as an up-front fixed fee will increase the incentive for international JV as it will help the foreign firm to raise its profit through fixed-fee or output royalty if the restriction on foreign shareholding restricts the foreign firm to increase its profit. In a paper Svejnar and Smith (1984) have shown that equity share and transfer price can act as substitutes.

2. Then firm 1 decides whether to make another agreement with firm 3.[13] If firm 2 has previously rejected the offer of firm 1, firm 1 will give a take-it-or-leave-it JV offer to firm 3. Upon acceptance or rejection by firm 3, firm 1 will decide whether to give an offer to firm 2. So, like licensing, we consider a similar extensive form game under JV.

Before going further, it is worth discussing the decision making procedure in the JV, which is always an important concern to the JV partners and important for market outcomes. There are different ways of modeling the decision-making in a JV. To make our analysis simple, we assume that in the JV the foreign partner (firm 1) gives the technical know-how and the JV management takes the production decision to maximize the aggregate profit of the JV firm, as considered in Kwoka (1992). Since we consider that the JV partners get only dividend incomes according to their equity shares in the JV and there is no moral hazard problem, our results remain unaffected even if we consider that the host-country firm takes the production decision in the JV (see, e.g., Purakayastha, 1993). However, the incentive for JV formation will increase if we consider that firm 1 takes the production decision in the JV since, by making two JVs, firm 1 can achieve collusion in the product market. We are not going to discuss this possibility here since it will not add much insight to our main analysis.

Proposition 2: *(a) If $\overline{\alpha} > \alpha^*$, firm 1 will form a JV with firm 2 and firm 1 will transfer its technology to the JV only.*

(b) If $\overline{\alpha} \leq \alpha^*$, firm 1 will prefer licensing to both firm 2 and firm 3 compared to a JV with firm 2.[14]

Proof: If firm 1 has already formed a JV with one of the host-country firms and wants to transfer its technology to the other firm, in our structure of JV, it is better for firm 1 to transfer its technology to the other firm through licensing than forming another JV.[15]

If firm 1 and firm 2 form a JV and firm 1 makes another licensing contract with firm 3, the profit of firm 1 is

$$\alpha\pi_J(c,c)+\pi_3(c,c), \tag{2}$$

where $\pi_J(c,c)$ shows the duopoly profit of the JV firm.

---

[13] Often the JV agreement prevents a firm to make another collaboration in the same market In our framework, such restriction will provide further incentive for JV. However, we rule out this possibility here. Often it is found that a firm is producing similar products in a market where it has a JV already (see, e.g., Mukherjee and Sengupta, 2001).

[14] If $\overline{\alpha} = \alpha^*$, we assume that firm 1 does licensing.

[15] If we allow lump-sum payment in the JV then firm 1 is indifferent between licensing and forming a JV with a host-country firm, given that it has already formed a JV with the other host-country firm.

On the other hand, if firm 1 and firm 2 form a JV and firm 1 does not transfer its technology to firm 3, the profit of firm 1 is $\alpha\pi_J(c)$, where $\pi_J(c)$ shows the monopoly profit of the JV firm.

Therefore, given a JV between firm 1 and firm 2, firm 1 does not have the incentive to license its technology to firm 3 provided

$$\alpha\pi_J(c) > \alpha\pi_J(c,c) + \pi_3(c,c) \Rightarrow \alpha > \frac{\pi_3(c,c)}{(\pi_J(c) - \pi_J(c,c))} \; (= \alpha^c, say).^{16} \quad (3)$$

Note that $\alpha^c < 1$ provided

$$\pi_J(c) > \pi_J(c,c) + \pi_3(c,c), \quad (4)$$

i.e., if monopoly profit is greater than the duopoly industry profit, which is always true under our demand specification.

However, firm 1 prefers this JV compared to licensing provided

$$\alpha\pi_J(c) > 2\pi(c,c) \Rightarrow \alpha > \frac{2\pi(c,c)}{\pi_J(c)} \; (= \alpha^*, say)^{17} \quad (5)$$

and $\alpha^* < 1$ provided

$$\pi_J(c) > 2\pi(c,c). \quad (6)$$

While constructing (5), we have assumed that when making a JV, firm 1 is not giving its technology to firm 3. This is always consistent with condition (3) since $\alpha^* > \alpha^c$.

Since firm 1 gives a take-it-or-leave-it offer to the host-country firm under JV, firm 1 holds $\overline{\alpha}$ proportion of shares in a JV.

(a) The above argument shows that when $\overline{\alpha} > \alpha^*$, firm 1 offers a JV contract to firm 2 with firm 1's proportion of shareholding as $\bar{\alpha}$ and transfers its technology to the JV only. So, it is always optimal for the firm 2 to accept this JV contract.

(b) Condition (5) shows that if $\alpha \leq \overline{\alpha}$, it is not possible for firm 1 to be better-off under JV compared to licensing. Hence, in this situation, firm 1 licenses its technology to both host-country firms. Q.E.D.

In the JV, firm 1 receives dividend income after production. So, if firm 1's shareholding in the JV is sufficiently large, it does not want to create another source

---

[16] It is easy to understand that if we have considered foreign control of the JV, it would affect the value of α^c by affecting $\pi_J(c,c)$ and $\pi_3(c,c)$.

[17] Note that $\pi_J(c,c) = \pi_3(c,c) = \pi(c,c)$.

of production by transferring its technology to firm 3 when it has formed a JV with firm 2. Thus, JV generates higher industry profit by creating a monopoly of the JV firm, while the market becomes duopoly under licensing. Since monopoly profit is greater than the duopoly industry profit, firm 1 is better-off under JV with firm 2 compared to licensing the technology to both host-country firms when firm 1 holds sufficiently large amount of shares in the JV. Thus, a suitably designed JV can guarantee exclusive technology transfer by eliminating the commitment problem faced by firm 1.

Thus, we show that when firm 1 cannot commit to an exclusive licensing contract, a JV between firm 1 and a host-country firm solves this commitment problem if the government regulated foreign shareholding is sufficiently large.[18] But, if the government regulated foreign shareholding is sufficiently small, firm 1 is not able to extract more profit under JV compared to licensing and does not form JV. However, it is clear that if firm 1 could use fixed-fee or royalty as a substitute for dividend incomes (see, Svejnar and Smith, 1984), JV is always a preferable option to firm 1.

At this stage, it is worth mentioning the difference between the JV agreement considered in this paper and a licensing scheme with profit sharing arrangement. In case of licensing with profit sharing arrangement, firm 1 is an outsider and does not have role in decision making at the firm 2. Hence, it can impose a constraint on firm 1's earning under licensing ex-post production if there is opportunistic behavior of firm 2. For example, firm 2 can imitate the foreign technology after getting it and can avoid ex-post payment by producing with the imitated technology (see, e.g., Gallini and Wright, 1990 and Rockett, 1990). On the other hand, by making a suitable combination of voting and non-voting shares, firm 1 can get significant power in the management of the JV firm (see, e.g., Dasgupta and Tao, 1998). Hence, a suitably formed JV also helps firm 1 to eliminate the opportunistic behavior of firm 2. Therefore, JV not only makes the payment ex-post production, JV also helps the firm 1 to receive payment ex-post production by providing managerial control in the JV firm. This possibility of having the managerial control in the JV makes the JV agreement significantly different from a licensing contract with profit sharing arrangement.

2.2 Choice of foreign technology

So far we have considered the problem under licensing and JV, for a given foreign technology and a given proportion of government regulated foreign shareholding.

It is well documented in the literature that international technology transfer involves significant amount of cost to be incurred by the technology supplier and higher amount of cost may be required for better technology (see, e.g., Teece, 1977,

---

[18] In a different context, viz., in case of vertically separated industry, Hart and Tirole (1990) have shown that vertical merger can solve the commitment problem when the exclusive-dealing contracts between the upstream firm and downstream firms are not enforceable.

1981 and Lee and Shy, 1992).[19] Hence, both the quality of the technology to be transferred and the mode of the technology transfer become the important questions faced by the foreign technology supplier.

Assume that firm 1 can transfer a technology corresponding to the marginal cost of production c but at a cost $f(c)$, where $c \in [0, a]$. Lower marginal cost comes with greater expenses on the improved technology and we assume that $f' < 0, f'' > 0$, $f''' = 0$ and $f(a) = 0$. While deciding the quality of the transferred technology, firm 1, however, internalizes its following action regarding the mode of entry.

Proposition 3: *Assume that* $\frac{1}{2} < f''(c)$.

(a) If the government regulated maximum shareholding of firm 1 is sufficiently small (i.e., $\overline{\alpha} \leq \frac{8}{9}$*), a rise in firm 1's maximum shareholding over* $[0, \frac{8}{9}]$ *does not affect the quality of the technology to be transferred in the host-country.*

(b) If the government regulated maximum shareholding of firm 1 is not sufficiently small (i.e., $\overline{\alpha} > \frac{8}{9}$*), firm 1 transfers better technology compared to the optimal technology under case (a) of this proposition. Any further increment in firm 1's maximum shareholding improves the quality of the transferred technology.*

Proof: First, consider the situation where the firm 1 wants to transfer its technology through licensing ex-post technology choice. The analysis of the previous section shows that, given the quality of the technology, firm 1 gets $2\pi(c, c)$ under licensing. Therefore, if firm 1 licenses its technology ex-post technology choice, it decides the quality of the technology by maximizing the following expression:

$$\underset{c}{Max}\, 2\pi(c, c) - f(c). \tag{7}$$

So, the optimal quality of the technology satisfies the following condition:

$$2\pi'(c, c) - f'(c) = 0. \tag{8}$$

Given the demand specification, condition (8) reduces to

$$-\frac{4(a - c)}{9} - f'(c) = 0. \tag{8'}$$

Second order condition for this maximization problem requires that $\frac{4}{9} < f''(c)$ and it is satisfied by assumption. Let c_l^* be the optimal quality of the technology that satisfies the equation (8'). So, the payoff of firm 1 ex-post technology transfer is $2\pi(c_l^*, c_l^*)$.

---

[19] We assume that the foreign firm needs to incur the cost of technology transfer even if it opens a fully owned subsidiary in the host-country. This is possible if we consider that the foreign technology needs to make compatible for the host-country environment. Or, we can think that this cost refers to the opportunity cost of technology transfer.

However, given the quality of the technology c_l^*, firm 1 does licensing ex-post technology choice if $\overline{\alpha} \leq \alpha^*(c_l^*)$, where $\alpha^*(c_l^*) = \frac{2\pi(c_l^* c_l^*)}{\pi(c_l^*)}$. Given the demand specification, we find that $\alpha^*(c_l^*) = \frac{8}{9}$.[20]

(a) So, it is optimal for firm 1 to license ex-post technology choice if $\overline{\alpha} \leq \frac{8}{9}$. Since (8) or (8') does not depend on $\overline{\alpha}$, any change in $\overline{\alpha}$ between 0 and $\frac{8}{9}$ does not affect c_l^*.

(b) If $\overline{\alpha} > \frac{8}{9}$, firm 1 will do JV ex-post technology choice. So, in this situation, while deciding the quality of the transferred technology, firm 1 will maximize the following expression instead of expression (7):

$$\underset{c}{Max}\, \overline{\alpha}\pi_J(c) - f(c). \tag{9}$$

The first order condition for maximization gives

$$\overline{\alpha}\pi_J'(c) - f'(c) = 0. \tag{10}$$

Given the demand specification, condition (10) reduces to

$$-\frac{2\overline{\alpha}(a-c)}{4} - f'(c) = 0. \tag{10'}$$

Second order condition for this maximization problem requires that $\frac{\overline{\alpha}}{2} < f''(c)$. Since, $\overline{\alpha}$ can take the maximum value of 1, we assume that $\frac{1}{2} < f''(c)$. Hence, the second order condition for maximization holds $\forall \alpha \in (\frac{8}{9}, 1]$.

Let c_J^* shows the quality of the technology that solves the equation (10'). It is easy to check from (8') and (10') that $c_J^* < c_l^*$ since firm 1 does JV for $\overline{\alpha} > \frac{8}{9}$. This implies that if firm 1 understands that ex-post technology choice it will form a JV, it will develop relatively better technology compared to the situation where it will license its technology ex-post technology choice.

Further, we find from (10') that as $\overline{\alpha}$ increases, it reduces c_J^*, since

---

[20] Note that $\alpha^*(c_l^*)$ does not depend on the marginal cost of production can be true even for non-linear market demand. For example, if the demand function is $P = a - q^\delta$, with $\delta > 1$, we find that $\alpha^*(c_l^*) = 2^{\frac{1}{\delta}} \left(\frac{1+\delta}{2+\delta}\right)^{\frac{(1+\delta)}{\delta}}$ and does not depend on the marginal cost of production. However, the value of $\alpha^*(c_l^*)$ depends on the assumption that there are two host-country firms producing under licensing. Generally, the value of $\alpha^*(c_l^*)$ is $\frac{4n}{(n+1)^2}$ if n firms produce under licensing.

$\frac{\partial c_J^*}{\partial \bar{\alpha}} = -\frac{(a-c)}{(2f''(c)-\bar{\alpha})} < 0$. Q.E.D.

Sufficiently higher foreign shareholding (i.e., $\bar{\alpha} > \frac{8}{9}$) has following effects. Firstly, it increases concentration in the host-country market compared to the situation where $\bar{\alpha} \leq \frac{8}{9}$. Secondly, any increment in the foreign shareholding over $(\frac{8}{9},1]$ reduces the profit of the host-country firm for any given quality of the transferred technology. Lastly, a higher shareholding of firm 1 over $(\frac{8}{9},1]$ encourages firm 1 to transfer relatively better quality technology to the host-country. Therefore, while higher shareholding over $(\frac{8}{9},1]$ increases efficiency in production due to the transfer of a relatively better technology, higher foreign shareholding increases market concentration and also reduces the host-country firm's share of profit. So, the net effect on host-country welfare is ambiguous and we will discuss this issue in the next subsection.

The above proposition shows that firm 1 will transfer better quality technology under JV compared to licensing. This holds under the assumption that the cost of technology transfer is positively related to the quality of the technology and the second order condition for maximization holds. However, the following proposition, where, for simplicity, we consider an extreme situation of same cost of technology transfer irrespective of the quality of the technology, shows that this is not necessarily the case.

Proposition 4: *Suppose.* $f(c) = k$ *for all* $c \in [0,a]$ *and* $k < 2\pi(0,0)$. *Firm 1 will transfer best technology under JV and licensing.*

Proof: We know that firm 1 will do JV provided it receives higher profit under JV compared to licensing. So, given the assumption $k < 2\pi(0,0)$, it is clear from (8) and (10) that firm 1 will always transfer the best technology under both JV and licensing, since, now we have $f'(c) = 0$. Q.E.D.

2.3 Host-country policy

We have so far focused on a given maximum proportion of shareholding of firm 1. In this section, we will examine the optimal proportion of the foreign shareholding that will maximize the host-country welfare. That is we will solve the first stage of our game where the host-country government takes the decision.

Let us first consider the situation where the firm 1 transfers same technology under JV and licensing (Proposition 4).

Proposition 5: *Assume that* $f(c) = k$ *for all* $c \in [0,a]$, $k < 2\pi(0,0)$ *and* $\bar{\alpha} > \alpha^* = \frac{8}{9}$. *Welfare of the host-country is higher under licensing compared to JV.*

Proof: Given the assumption of this proposition, we find from Proposition 4 that the firm 1 transfers best technology under both JV and licensing.

If firm 1 forms a JV, the host-country welfare is

$$(1-\bar{\alpha})\pi_J(0)+CS(0). \tag{11}$$

But, if the firm 1 licenses its technology to the host-country firms, the host-country welfare is $CS(0,0)$.

It is easy to check that we have

$$CS(0,0)+2\pi(0,0)>CS(0)+\pi(0) \tag{12}$$

$$\text{Or, } CS(0,0)>CS(0)+(1-\bar{\alpha})\pi_J(0)+[\bar{\alpha}\pi_J(0)-2\pi(0,0)]. \tag{12'}$$

The condition for profitable JV implies that $\bar{\alpha}\pi_J(0)>2\pi(0,0)$. So, it is clear from (12') that $CS(0,0)>CS(0)+(1-\bar{\alpha})\pi_J(0)$ and the host-country welfare is higher under licensing compared to JV. Q.E.D.

Proposition 5 shows that while JV is privately profitable to the collaborating firms, host-country welfare is higher under licensing. Hence, there is a conflict between the interests of the collaborative firms and the host-country.

Our above result contradicts the finding of Tang and Yu (1990), where JV generates higher host-country welfare compared to technology licensing. We have found that if JV prevents multiple technology transfer, it reduces welfare of the host-country compared to licensing. It is easy to understand that if firm 1 has other possible ways of extracting profits (e.g., an up-front fixed fee) under JV, it will further reduce the welfare of the host-country.

It must be noted that while constructing the welfare function of the host-country, we have given same weight to the consumer surplus and the profit share of the domestic firm. Depending on the government's ideology and the product identity (e.g., necessary or luxury products), these weights may differ. If consumer surplus gets more weight in the welfare function then JV makes things further worse-off; however, on the other hand, if the profit of the domestic firm is the main concern to the host-country then JV tends to increase the welfare of the host-country.

In Proposition 5 we have assumed same cost of technology transfer for all types of technology. Now, we relax this assumption and, like Proposition 3, we assume that better technology requires higher cost of technology transfer and $\frac{1}{2}<f''(c)$.

It follows from Proposition 3 that the host-country welfare is $W_l=CS(c_l^*,c_l^*)$ for any $\bar{\alpha}<\frac{8}{9}$. But, for $\bar{\alpha}>\frac{8}{9}$, host-country welfare is $W_J=(1-\bar{\alpha})\pi_J(c_J^*)+CS(c_J^*)$, where $c_J^*=c_J^{mm}$ for $\bar{\alpha}=1$. Realizing the relationship between the optimal foreign technology and the maximum foreign shareholding, we find that $\frac{\partial W_J}{\partial \bar{\alpha}} \gtrless 0$ as

$$\frac{3}{2} \substack{\geq \\ <} \frac{(f''(c)-\bar{\alpha})}{(1-\bar{\alpha})}. \tag{13}$$

Second order condition for choosing the optimal foreign technology provides the restriction that $f''(c) > \frac{1}{2}$. It is easy to check that, given $f'''(c) = 0$, the right hand side (RHS) of (13) is increasing (decreasing) in $\bar{\alpha}$ for $f''(c) > (<) 1$.

Proposition 6: *(a) Suppose, $\frac{1}{2} < f''(c) < 1$. In this situation, the host-country will either allow very low foreign shareholding such that licensing occurs, i.e., $\alpha \in [0, \frac{8}{9}]$, or will allow fully owned subsidiary of the foreign firm, i.e., $\bar{\alpha} = 1$.*

(b) Suppose, $1 < f''(c)$. Here, the host-country will not allow fully owned subsidiary of the foreign firm. The host-country will either allow very low foreign shareholding such that licensing occurs, i.e., $\alpha \in [0, \frac{8}{9}]$, or will allow foreign shareholding of $(\frac{8}{9} + \varepsilon)$ (for $(9f''(c) - 8) > \frac{3}{2}$) or $\bar{\alpha}_J \in (\frac{8}{9}, 1)$ (for $(9f''(c) - 8) < \frac{3}{2}$).

Proof: (a) If $\frac{1}{2} < f''(c) < 1$, RHS of (13) is decreasing in $\bar{\alpha}$ and is less than the left hand side (LHS) of (13) for $\bar{\alpha} = \frac{8}{9}$. So, $\frac{\partial W_J}{\partial \alpha} > 0$ for any $\alpha \in (\frac{8}{9}, 1]$. Therefore, if $CS(c_l^*, c_l^*) > (<) CS(c_J^{mm})$, the host-country welfare maximizing α is between $[0, \frac{8}{9}]$ (equal to 1).

(b) If $1 < f''(c)$, RHS of (13) is greater than the LHS of (13) at $\bar{\alpha} = 1$ and hence, $\frac{\partial W_J}{\partial \alpha} < 0$ at $\bar{\alpha} = 1$. But, for $\bar{\alpha} = \frac{8}{9}$, RHS of (13) is less than, equal to or greater than the LHS of (13) provided $(9f''(c) - 8) \substack{\leq \\ >} \frac{3}{2}$. So, at $\bar{\alpha} = \frac{8}{9}$, $\frac{\partial W_J}{\partial \alpha} \substack{\geq \\ <} 0$ as $(9f''(c) - 8) \substack{\leq \\ >} \frac{3}{2}$.

Suppose $f''(c)$ is such that $\frac{\partial W_J}{\partial \alpha} > 0$ at $\bar{\alpha} = \frac{8}{9}$. Hence, in this situation, W_J attains a maximum value at, say, $\bar{\alpha}_J$, where $\bar{\alpha}_J \in (\frac{8}{9}, 1)$. Therefore, the optimal $\alpha \in [0, \frac{8}{9}]$ if $CS(c_l^*, c_l^*) > W_J(\bar{\alpha}_J)$ and the foreign firm does licensing. But, for $CS(c_l^*, c_l^*) < W_J(\bar{\alpha}_J)$, the optimal $\bar{\alpha} = \bar{\alpha}_J$ and the foreign firm does JV with foreign shareholding equals to $\bar{\alpha}_J$.

Next, consider that $f''(c)$ is such that $\frac{\partial W_J}{\partial \alpha} < 0$ at $\bar{\alpha} = \frac{8}{9}$. Hence, in this situation, optimal $\alpha \in [0, \frac{8}{9}]$ for $CS(c_l^*, c_l^*) > (1 - \frac{8}{9})\pi_J(c_J^*(\frac{8}{9})) + CS(c_J^*(\frac{8}{9}))$ and the foreign firm does licensing. But, if $CS(c_l^*, c_l^*) < (1-\frac{8}{9})\pi_J(c_J^*(\frac{8}{9})) + CS(c_J^*(\frac{8}{9}))$, the optimal $\overline{\alpha} = \frac{8}{9} + \varepsilon$, where $\varepsilon > 0$ and very small, and hence, the foreign firm does a JV with its shareholding equals to $(\frac{8}{9} + \varepsilon)$. Q.E.D.

The reason for the above finding is as follows. Sufficiently higher foreign shareholding has two effects on the host-country welfare. On one hand, it helps to attract relatively better technology to the host-country but, on the other hand, it reduces host-country welfare by reducing the profit of the host-country firm for a given quality of the technology. We have seen in proposition 3 that $\frac{\partial c_J^*}{\partial \overline{\alpha}} = -\frac{(a-c)}{(2f_I''(c)-\overline{\alpha})}$. Hence, the effect of higher $\overline{\alpha}$ on c_J^* reduces with lower market size of the host-country, i.e., a, and/or higher rate of change of the marginal cost of technology choice, i.e., $f''(c)$. If the size of the host-country market is sufficiently small, the effect of higher foreign shareholding on the quality of technology is not significant and hence, higher government regulated foreign shareholding reduces welfare of the host-country. In this situation, host-country will not allow sufficiently higher foreign shareholding and as a result, licensing will take place.

Higher shareholding of firm 1 attracts sufficiently better technology when the market size of the host-country is not very small and it is not very costly to transfer better technology to the host-country (i.e., when $\frac{1}{2} < f''(c) < 1$). In this situation, it is optimal for the host-country to allow fully owned subsidiary of the foreign firm.

Now, consider that the host-country market is not sufficiently small but it is sufficiently costly to transfer better technology to the host-country (i.e., $f''(c) > 1$).Then higher foreign shareholding attracts relatively better technology to the host-country but not as good as like the previous situation (i.e., where $\frac{1}{2} < f''(c) < 1$). Hence, in this situation, after certain amount of foreign shareholding, the loss of profit of the host-country firm dominates the gain from relatively better foreign technology. Therefore, after a certain amount of foreign shareholding, any rise in the foreign shareholding reduces the host-country welfare. So, the host-country welfare is maximized if it allows JV with this critical foreign shareholding, i.e., up to $\overline{\alpha}_J$.

Before concluding the paper, it is worth mentioning the importance of the committed host-country policy for our results. It is evident from our analysis that ex-post technology choice, welfare of the host-country is higher under licensing compared to JV. Therefore, after stage 2 of our game, the government of the host-country has an incentive for not allowing the JV. However, if there is a chance of policy reversal after stage 2, the firm 1 has less incentive for transferring a relatively better technology to the host-country. So, unlike Al-Saadon and Das (1996), this paper shows that the host-country can enjoy higher welfare under a committed government policy.

3. Conclusion

Last few decades have seen significant changes in the policies of the developing countries. Developing countries are liberalizing their policies to attract better technologies from the technologically advanced firms. However, initially the liberalizing countries were mainly interested for technology licensing agreements. The cases for international JVs are increasing dramatically in the recent years.

This paper examines the optimal host-country policy regarding foreign investment when technology licensing involves opportunistic behavior of the foreign firm. The formation of a JV may reduce the number of technology transfers and helps to increase market-concentration. Hence, a host-country can have lower welfare under JV than licensing while the firms are better-off under JV compared to licensing. However, higher market-concentration can increase efficiency by encouraging the foreign firm to transfer relatively better technology to the host-country. If this efficiency-gain is sufficiently large then the JV can make both the firms and the host-country better-off compared to technology licensing. For sufficiently large efficiency-gain, the host-country is better-off by allowing fully owned subsidiary of the foreign firm.

References

Al-Saadon, Y. and S. P. Das, 1996 : 'Host-country policy, transfer pricing and ownership distribution in international joint venture: a theoretical analysis', *International Journal of Industrial Organization 14: 345-64.*

Bagchi, A. K., 1987 : 'Public intervention and industrial restructuring in China, India and Republic of Korea', *ARTEP, ILO, New Delhi.*

Das, S. P., 1998 : 'On the choice of international joint venture: the role of policy moral hazard', *Journal of Policy Reform 2: 135-50.*

Dasgupta, S. and Z. Tao, 1998 : 'Contractual incompleteness and the optimality of equity joint ventures', *Journal of Economic Behavior and Organization 37: 391-413.*

Economic and Political Weekly, 1985, *Special Issue.*

Falvey, R. E. and H. O. Fried, 1986 : 'National ownership requirements and transfer pricing', *Journal of Development Economics 24: 249-54.*

Gallini, N. T. and B. D. Wright, 1990 : 'Technology transfer under asymmetric information', *RAND Journal of Economics 21: 147-60.*

Hart, O. and J. Tirole, 1990 : 'Vertical integration and market foreclosure', *Brookings Papers on Economic Activity, 205 - 76 and 285 - 86.*

Kabiraj, T. and S. Marjit, 1992 : 'Transfer or not to transfer the best technology under threat of entry - the case of price competition', in B. Dutta et al., ed., *Game Theory and Economic Applications, Springer-Verlag, Berlin: 356-68.*

Kabiraj, T. and S., Marjit, 1993 : 'International technology transfer under potential threat of entry - a Cournot-Nash framework', *Journal of Development Economics 42:75-88.*

Katz, M. and C. Shapiro, 1985 : 'On the licensing of innovations', *RAND Journal of Economics 16: 504-20.*

Korbin, S. J., 1998 : 'Ownership of American manufacturing subsidiaries in developing countries', *Management International Review, Special Issue: 73-84.*

Kwoka, J. E., Jr., 1992 : 'The output and profit effects of horizontal joint ventures', *Journal of Industrial Economics, 40: 325 - 38.*

Lee, F. C. and O. Shy, 1992 : 'A welfare evaluation of technology transfer to joint ventures in the developing countries', *The International Trade Journal 7: 205-20.*

Marin, D. and M. Schnitzer, 1995 : 'Tying trade flows: a theory of countertrade with evidence', *American Economic Review 85: 1047-64.*

Marjit, S., 1990a : 'On a non-cooperative theory of technology transfer', *Economics Letters 33: 293-98.*

Marjit, S., 1990b : 'Rationalizing public-private joint venture in an open economy', *Journal of Development Economics 33: 377-83.*

Marjit, S. and A. Mukherjee, 1998 : 'Technology collaboration and foreign equity participation: a theoretical analysis', *Review of International Economics 6: 142-50.*

Markusen, J. R., 1995 : 'The boundaries of multinational enterprises and the theory of international trade', *Journal of Economic Perspectives 9: 169-89.*

Mukherjee, A., 2000 : 'Host-country policy – commitment or no-commitment: a theoretical analysis', *Journal of Policy Reform, 4: 75-89.*

Mukherjee, A., 2001 : 'Technology transfer with commitment', *Economic Theory, 17: 345 – 69.*

Mukherjee, A., 2002 : 'Subsidy and entry: the role of licensing', *Oxford Economic Papers, 54: 160 – 71.*

Mukherjee, A. and S. Sengupta, 2001 : 'Joint ventures versus fully owned subsidiaries: multinational strategies in liberalizing economies', *Review of International Economics, 9: 163 - 80.*

Pack, H. and K. Saggi, 1997 : 'Inflows of foreign technology and indigenous technological development', *Review of Development Economics 1: 81-98.*

Purakayastha, D., 1993 : 'Firm-specific advantages, multinational joint ventures and host-country tariff policy', *Southern Economic Journal 60: 89-95.*

Rockett, K., 1990 : 'The quality of licensed technology', *International Journal of Industrial Organization 8: 559 - 574.*

Saqib, M., 1995 : 'An overview of the economic reform policy and FDI in India with special reference to the USA', in S. P. Gupta et al, ed., *Prospects of Foreign Direct Investment in India in Post Liberalisation Era, Indian Council for Research on International Economic Relations: 81-99.*

Singh, N., 1992 : 'Multinationals, technology and government policy', K. Basu and P. Nayak, ed., *Development Policy and Economic Theory, Oxford University Press, New Delhi.*

Sinha, U. B., 2001a : 'Imitative innovation and international joint ventures: a dynamic analysis', *International Journal of Industrial Organization, 10: 1527 – 62.*

Sinha, U. B., 2001b : 'International joint venture, licensing and buy-out under asymmetric information', *Journal of Development Economics, 66: 127 – 51.*

Svejnar, J. and S. C. Smith, 1984 : 'The economics of joint venture in less developed countries', *Quarterly Journal of Economics 119: 149-67.*

Teece, D. J., 1977 : 'Technology transfer by multinational firms: the resource cost of transferring technological know-how', *Economic Journal 87: 242-61.*

Teece, D. J., 1981 : 'The market for know-how and the efficient international transfer of technology', *ANNALS, AAPSS, November: 81-95.*

Tang, M. J. and C. M. J. Yu, 1990 : 'Foreign market entry: production-related strategies', *Management Science 36: 476-89.*

Arthaniti 2 (1-2), 97-107 / 2003

Child Labour and Economic Development : A Theoretical Analysis

Manash Ranjan Gupta*

Abstract

In this paper, we develop a four sector static model of a small open economy with special consideration to the child labour market. The model is used to analyse the general equilibrium effects of capital accumulation on the child labour market and on national income.

JEL Classification : F10, J10, J13, I28.

Keywords : Child labour, General equilibrium, Foreign capital, Informal Sector.

---

1. Introduction

The problem of child labour is widespread at present in the less developed countries. Even the economically advanced countries of today had to face this problem at the early stage of their development. The rate of decline of the participation rates for children aged between 10 to 14 years, during the late 20th Century is marginal in many of the African and Asian countries. In India and China, these rates were 35.43% and 47.85% in 1950; and, in 1990, they declined to 16.68% and 15.24% only. In Ethiopia, the participation rate was 52.95% in 1950 and declined to only 43.47% in 1990.

The awareness of and concern for child labourers have increased throughout the world; and there is a strong feelings that the institution of child labour must be abolished. One simple way is to ban it; and across the world, there is a strong opinion in favour of a ban. For example, there is a view to impose trade sanctions on the import of those products which are produced using child labour. International Labour Organizations also favour at least partial ban on child labour and the various restrictions on their employers. There are suggestions for labelling "child labour free" bands on the products so that the consumers can confine their consumption to such goods. In U.K., many N.G.O.s are involved in the naming and shaming of stores which sell products produced with the use of child labour.

There exists a small theoretical literature on the economics of child labour market. Basu and Van (1998) analyse the implication of imposing a ban on child labour in a static general equilibrium model which shows the possibility of a child labour trap. Swinnerton and Rogers (1999) shows that the child labour trap can not exist in Basu

---

* Economic Research Unit, Indian Statistical Institute, Kolkata - 700 108, West Bengal, India.

and Van (1998) model if the non-labour income is equally distributed. Ranjan (1999) explains the existence of child labour in terms of credit market imperfection and shows that a ban on child labour may reduce the welfare of the households who send their children in labour market.Jafarey and Lahiri (2002) analyse the effects of trade snactions in reducing the child labour and show how the nature of the credit market is related to these effects. Gupta (2000) shows how the wage rate of a child worker is determined by the bargaining power of his employer and his guardian. Gupta (2002) also analyses the effect of trade sanctions on the unemployment in the adult labour market and on the size of the child labour market.

In this paper, we develop a theoretical model of a dual economy with special consideration to the interaction between the adult labour market and the child labour market. Two types of labour are considered in this model : adult labour and child labour. There are four sectors in the economy. Urban formal sector, the urban informal sector and the rural sector employ adult labour. Also there is child labour sector. The model is used to study the general equilibrium effects of economic expansion (capital accumulation) on the wage rate of the child workers and on the size of the child labour market. We get some interesting results from this excercise. The increase in capital stock lowers the size of the child labour market and raises the wage rate there. This also lowers the national income when the demand for child labour using product is price elastic.

The paper is organised as follows : Section 2 describes the basic model and its working. The comparative static analysis is also described in section 3. Concluding remarks are made in section 4.

2. The Model

We consider a small open economy consisting of two sectors :- adult labour sector and child labour sector. The adult labour sector is further subdivided into three sub sectors : urban formal sector, urban informal sector and the rural sector. The products of all the adult labour using sectors are internationally traded; and hence their prices are exogenously given. However, child labour using sector may produce either traded or non-traded goods. The price of the non-traded product is determined by the interaction between the demand and supply forces within the economy. The rural sector, the urban formal sector and the urban informal sector use adult labour and capital as inputs. The production functions of these sectors exhibit CRS and the marginal productivity of each input is positive and diminishing. The child labour sector uses child labour as the only input and the production function of this sector is subject to the law of diminishing returns.

We assume that the endowment of capital and adult labour are given and these two inputs are fully employed. We consider perfect mobility of capital among all the three adult labour using sectors. So we have a common rate of return on capital in all the adult labour using sectors.

The adult workers in the informal sector and rural sector send their children to work and the supply function of child labour is derived from their rational behaviour.[1] The supply of child labour is assumed to depend positively on the child wage rate and on the size of the adult labour force in the informal and rural sector and negatively on the adult wage rate in these two sectors. The decision of supplying the children to work is made by their parents. In equilibrium, the supply-demand equality is ensured in the child labour market because the wage rate of child labour is flexible.

Urban formal sector's wage rate is institutionally fixed but wage rate in all other adult labour using sectors are flexible.[2] So, unemployment does not exist in the adult labour market.

Adult labourers migrate from the rural to the urban sector and are absorbed in the urban formal sector or in the urban informal sector. The migration mechanism is of Harris-Todaro (1970) type and, in migration equilibrium, the actual rural wage and the expected urban wage are equalised.

All the markets are assumed to be perfectly competitive. The profit maximising equilibrium of competitive firms and the CRS property of the production functions lead to the equality between price and unit cost in each of the adult labour using sectors. However, in the child labour using sector, price is equal to the marginal cost in equilibrium; and the marginal cost is a positive function of the wage rate and the level of output in that sector.

It is assumed that the urban sector is more capital intensive than the rural sector. Here, the urban capital intensity is defined as the ratio of capital available to the urban sector to the adult urban labour force.

Following notations are used in this model :

j = u, r, i, C.

u = Urban formal sector.

r = Rural sector.

i = Urban informal sector.

C = Child labour sector.

---

[1] Adult worker is assumed to maximize utility which is a positive function of income - sum of his own wage income and that of his children - and is a negative function of labour of his children.

[2] See Chandra and Khan (1993), Grinols (1991), Gupta (1997) etc. where similar assumptions have been made.

X_j = Output produced in the jth sector.

L_j = Level of employment in the jth sector.

L_C^S = Supply of child labour.

w_j = Wage rate in the jth sector.

k_j = Capital intensity of the jth sector.

R = Common rental rate on capital.

R_j = Rental rate on capital in the jth sector.

L = Total endowment of adult labour in the entire economy.

K = Stock of capital of the entire economy.

P_j = Price of the jth good.

f_j = Intensive production function of the jth sector.

F_C = Production function of the child labour sector.

ψ_j = Unit cost of production in the jth sector.

ψ_c = Marginal cost of production of the child labour using sector.

λ = Probability of getting urban formal job of an adult worker.

Y = National income of the economy.

D = Demand for product produced by the child labour using sector.

We use the following equations to describe the model.

The intensive production functions of the three labour using sectors are given by

$$X_u = f_u(k_u)\, L_u \tag{1}$$

$$X_r = f_r(k_r)\, L_r \tag{2}$$

$$X_i = f_i(k_i)\, L_i \tag{3}$$

The production function of the child labour using sector is :

$$X_c = f_c(k_c)\, L_c \tag{4}$$

The production functions of these sectors satisfy the following properties :

$f_j'(.) > 0$ and $f_j''(.) < 0$ for j = u, r and i.

$F_c'(.) > 0$ and $F_c''(.) < 0$.

The assumptions of CRS production function and the profit maximization of the competitive firms imply that price is equal to the unit cost in equilibrium in each of the adult labour using sectors. Hence, we have the following three equations :

$$P_u = \psi_u(\bar{w}_u, R) \tag{5}$$

$$P_r = \psi_r(w_r, R) \qquad (6) \text{ and}$$

$$P_i = \psi_i(w_i, R) \tag{7}$$

In the child labour using sector, profit maximisation implies price to be equal to the marginal cost; and the marginal cost is increasing in output and is higher than the average cost. Thus we get the competitive equilibrium equation for the child labour using sector as follows :

$$P_c = \psi_c(w_c, X_c) \tag{8}$$

satisfying $(\delta\psi_c / \delta w_c) > 0$ and $(\delta\psi_c / \delta X_c) > 0$.

If the child labour using sector produces a non-traded good, the demand function for this product is given by the following equation :

$$D = \dot{D}(P_c) \text{ with } D'(.) < 0 \tag{9}$$

Here, income effect is assumed to be nil for the sake of simplicity.

Equilibrium in that product market implies the equality between the demand for and supply of it. Thus, we have,

$$X_c = D(P_c) \tag{10}$$

Substituting equations (4) and (8) into equation (10), we get

$$F_c(L_c) = D(\psi_c(w_c, F_c(L_c))) \tag{10a}$$

Equations (10) and (10a) do not exist when it is a traded good.

The probability of getting urban formal job of the rural migrant is given by

$$\lambda = (L_u / (L_u + L_i)) \tag{11}$$

The Harris-Todaro (1970) migration equilibrium condition is given by :

$$w_r = \lambda \bar{w}_u + (1-\lambda) w_i \tag{12}$$

The supply function of the child labour is given by the following equation :

$$L_C^S = L_C^S(w_c, w_i, w_r, L_r+L_i) \tag{13}$$

where, $(\delta L_C^S / \delta w_c) > 0$, $(\delta L_C^S / \delta w_j) < 0$ and $(\delta L_C^S / \delta L_j) > 0$ for $j = i, r$.

As the child wage rate rises, the parents send more children to work. However, if wage rate of the adult worker rises, they send less children in the market. On the other hand, the supply of child labour varies positively with the number of the informal sector adult workers and the rural sector adult workers because these workers earn low wage income and send their children to work.

Child labour market is always in equilibrium.

Hence,

$$L_C^S = L_c \tag*{(14)}$$

The supply-demand balance equations for adult labour market and the capital market are given by the followings :

$$L_u + L_r + L_i = L \tag*{(15) and}$$

$$k_u L_u + k_r L_r + k_i L_i = K \tag*{(16).}$$

The optimum capital intensities of the three adult labour using sectors are functions of factor price ratios.

So, we have

$$k_u = k_u(\bar{w}_u / R) \text{ with } k_u'(.) > 0 \tag*{(17)}$$

$$k_r = k_r(w_r / R) \text{ with } k_r'(.) > 0 \tag*{(18), and}$$

$$k_i = k_i(w_i / R) \text{ with } k_i'(.) > 0 \tag*{(19)}$$

The national income of the economy is given by the following :

$$Y = \bar{w}_u L_u + w_r L_r + w_i L_i + P_c X_c + RK \tag*{(20)}$$

Using equations (9), (10), (12) and (17), we get

$$Y = w_r L + P_c X_c + RK \tag*{(20.1).}$$

Also we have the assumption that

$$\lambda k_u + (1-\lambda) k_i > k_r .$$

This implies that, on the average, the urban region is more capital intensive than the rural region.[3]

The working of the model is described as follows :

Equation (5) yields the equilibrium value of R, given P_u and $\bar{w}_u$. Then we get the values of w_r and w_i from equations (6) and (7), given P_r and P_i respectively. The values of k_u, k_r and k_i are obtained from equations (17), (18) and (19). Next, equation (12) yields λ, given the equilibrium values of w_r and w_i. Now, the values L_u, L_r and L_i can be obtained from equations (11), (15) and (16).

---

[3] See Chandra and Khan (1993), Corden and Findlay (1975) etc. for the interpretation of similar condition.

Substituting equation (4) into the equation (8) we get,

$$P_c = \psi_c (w_c \, . \, F_c (L_c)) \qquad (8a)$$

Equations (8a) and (14) along with (13) determine L_c and w_c, given P_c and already determined w_i, L_i, w_r and L_r.

The equilibrium values of X_u, X_r, X_i and X_c can be obtained from equations (1) to (4) respectively. Finally, equation (20.1) yields the equilibrium value of Y.

If the child labour using sector produces a traded good, then P_c is exogenous to the system and equation (10) does not exist. However, if this sector produces a non-traded good, then P_c is to be determined within this model. Using equations (8a) and (14) we have already solved L_c in terms of P_c. Then equations (10) and (4) solve for the equilibrium value of P_c.

3. Effects of Expansion

In this section, we analyse the effects of economic expansion on the working of the child labour market. Economic expansion means the increase in capital stock, K.

Note that, in this model, w_r, w_i, and R are determined independent of factor endowments. This is the indecomposable property of this model. So the increase in K does not affect w_r and w_i. Also the capital intensities – k_u, k_r and k_i remain independent of the change in K.

Now using equations (11), (15) and (16) we have

$$(L_u / \lambda) + L_r = L \qquad \text{(15a), and}$$

$$(k_u + ((1-\lambda) / \lambda)k_i) L_u + k_r \, . \, L_r = K \qquad \text{(16a).}$$

It can be easily shown that

$$(\delta L_u / \delta K) > 0 \text{ if } \lambda k_u + (1-\lambda) k_i > k_r.$$

So, if the urban region is more capital intensive than the rural region, then an expansion of capital stock raises the level of employment in the formal sector.[4] Hence $(L_r + L_i) = (L - L_u)$ falls in this case.

Now we use equations (13), (14) and (8a) and taking the total differential we get

$$\frac{dL_c}{dK} = (\delta L_c^s / \delta w_c) \cdot \left(\frac{dw_c}{dK} \right) + (\delta L_c^s / \delta (L_i + L_r)) \left(\frac{d(L_i + L_r)}{dK} \right);$$

and

$$0 = (\delta \psi_c / \delta w_c) . \left(\frac{dw_c}{dK} \right) + (\delta \psi_c / \delta X_c) . F_c'(L_c) . \left(\frac{dL_c}{dK} \right)$$

---

[4] See Chandra and Khan (1993) for derivation of a similar result.

Here,

$$\frac{dL_c}{dK} = \frac{-(\delta L_c^s / \delta(L_1 + L_r))(dL_u / dK)\cdot(\delta\psi_c / \delta w_c)}{[(\delta\psi_c / \delta X_c)F'(L_c).(\delta L_c^s / \delta w_c)] + (\delta\psi_c / \delta w_c)}$$

and

$$\frac{dw_c}{dK} = \frac{(\delta L_c^s / \delta(L_1 + L_r))(dL_u / dK)\cdot(\delta\psi_c / \delta X_c)F_c'(L_c)}{[(\delta\psi_c / \delta X_c)F'(L_c).(\delta L_c^s / \delta w_c)] + (\delta\psi_c / \delta w_c)}$$

Here $(dL_c / dK) < 0$ and $(dw_c / dK) > 0$. Hence we have the following proposition.

Proposition I : The increase in capital stock, K, lowers the size of the child labour market and raises the wage rate there.

Size of the child labour market and the wage rate there are determined by the interaction of the supply forces and demand forces. Income effect for the demand for child labour using product is assumed to be nil. Also child labour sector does not use capital as an input. So the capital accumulation does not have any effect on the demand for child labour. However, the size of the formal sector's labour force is increased because the urban region, by assumption, is more capital intensive. So $(L_1 + L_r)$ falls. Also the wage rates $-w_1$ and w_r- remain unchanged. As the workers in rural sector and informal sector send their children to work, the supply of child labour is reduced in this case. As a result, size of the child labour market is reduced and the equilibrium wage rate goes up.

If the child labour sector produces an inferior (normal) good, then capital accumulation raising total income lowers (raises) the demand for the product and hence lowers (raises) the demand for child labour. So the reduction in the size of the child labour market will be higher and the effect on the wage rate will be indeterminate in the case of an inferior good. However, in the case of a normal good, equilibrium wage rate is increased by a larger amount and the effect on the size of the child labour market is indeterminate.

The effect on national income, Y, is given by

$$(\delta Y/\delta K) = R + (\delta(P_c D_c) / \delta L_c)\ (\delta L_c/\delta K)$$

$$\text{or } (\delta Y/\delta K) = R + (\delta(P_c D_c) / \delta X_c)\ .\ F'\ (\delta L_c)\ .\ (\delta L_c/\delta K)$$

Here $(\delta Y/\delta K)$ may be negative if $(\delta(P_c D_c) / \delta X_c) > 0$; and this is possible if and only if the demand for child labour using product is price elastic.

If the economic expansion takes place through the inflow of foreign capital and if the entire foreign capital income is repatriated then national income is defined as

$$Y = RK_D + P_c X_c + w_c L_c\ ;$$

where $K_D(K_F)$ is exogenously given stock of domestic (foreign) capital; and

$$(\delta Y/\delta K_F) = (\delta(P_c D_c) / \delta X_c) . F'(L_c) . (\delta L_c/\delta K_F)$$

Then $(\delta Y/\delta K_F)$ is always negative. So we have the following proposition.

Proposition 2 : The economic expansion defined as the increase in capital-stock may be immiserizing if the demand for child labour using product is price-elastic and must be immiserizing when the additional capital stock is owned by the foreign country and its entire rental income is repatriated.

The rental rate on capital, R, and the wage rates in the adult labour market w_r and w_1 remain unaffected when the capital stock is increased. So the total wage income of the adult workers, $w_r . L$, remains unchanged. If the foreign capital stock is increased, then the rental income of domestic capitalists also remains unchanged. So the effect on national income is determined by the effect on income generated by the child labour sector, $P_c . X_c$.

As the size of the child labour market is reduced, total product of this sector also falls and hence price of that is increased. If the demand for the product is price-elastic, then $P_c X_c$ also falls. So national income is reduced. In the case of an expansion of domestic capital stock, this is not necessarily true because total rental income on domestic capital also goes up.

4. Conclusion

In this paper, we have analysed the working of a general equilibrium model of a dual economy which takes care of the interaction between the adult labour market and the child labour market. We have analysed the effect of economic expansion in the form of an increase in capital stock on the size of the child labour market. The size of the child labour market is reduced and the equilibrium wage-rate of a child worker is increased when economic expansion takes place. On the other hand this may produce a negative effect on national income through the loss of output in the child labour sector. However, this problem is not at all serious when child labour sector contributes insignificantly to the national income. In general, the contribution of the child labourers to the national income is very low; and the social marginal productivity (SMP) is zero if there exists unemployment in the adult labour market. This SMP may be negative when the existence of child labour market appears as an obstacle to the skill formation of the adult workers. On the other hand, the negative effect of economic expansion on the size of the child labour market appears to be interesting from the view point of policy analysis. It implies that the child labour problem can be reduced by adopting the policies favourable to economic growth; and the existence of child labour market is the result of economic backwardness - low level of capital stock. A ban on child labour to remove it is not necessary.

References

Addison, Tony, Bhalotra, Sonia, Coulter, Fiona and Heady, Christopher (1997) : 'Child labour in Pakistan and Ghana : A Comparative Study', *mimeo : University of Warwick, U.K.*

Ashagrie, Kebebew (1998) : 'Statistics on Child labour and Hazardous Child labour in Brief', *mimeo, Bureau of Labour Statistics, ILO, Geneva.*

Baland, J. M. and Robinson, J (2000) : 'Is Child Labour inefficient', *Journal of Political Economy, 108, 663-679.*

Basu, K. (1999) : 'Child Labour; Cause, Consequence and Cure', *Journal of Economic Literature, 37(3), 1083-1119.*

Basu, Kaushik and Van, Pham Hoang (1998) : 'The Economics of Child Labour', *American Economic Review, 88, 412-27.*

Bonnet, Michel (1993) : 'Child Labour in Africa', *International Labour Review, 132 (3), 371-89.*

Burra, Neera (1995) : 'Born to Work : Child Labour in India', *Oxford University Press, Delhi.*

Cain, Mead. T. (1977) : 'The Economic Activities of Children in a Village in Bangladesh', *Population and Development Review, 3(3), 201-27.*

Chandra, V. and Khan, M. A. (1993) : 'Foreign investment in the presence of Informal Sector', *Economica, 60, 79-103.*

Chandrasekhar, C. P. (1997) : 'The Economic Consequences of the Abolition of Child Labour', *Journal of Peasant Studies, 24(3), 137-79.*

Corden, W. M. and Findlay, R. (1975) : 'Urban unemployment, intersectoral capital mobility and development policy', *Economica, 42, 59-78.*

Cunningham, Hugh (1990) : 'The Employment and Unemployment of Children in England C. 1680-1851' *Past and Present, No. 126, 115-50.*

Eswaran, Mukesh (1996) : 'Fertility, Literacy and the Institution of Child Labour', *mimeo : University of British Columbia, Vancouver, Canada.*

Grinols, E. L. (1991) : 'Unemployment and Foreign Capital : The Relative Opportunity - Cost of Domestic Labour and Welfare', *Economica, 57, 107-121.*

Gupta, M. R. (1997) : 'Informal Sector and Informal Capital Market in a Small Open Less-developed Economy', *Journal of Development Economics, 52, 409-428.*

Gupta, M. R. (2000) : 'Wage Determination of a Child Worker : A Theoretical Analysis', *Review of Development Economics, 4 (2), 219-228.*

Gupta, M. R. (2002) : 'Trade Sanctions, Adult Unemployment and the supply of Child Labour : A Theoretical Analysis', *Development Policy Review, 20(3), 317-332.*

Gupta, M. R. (2002) : 'Child Labour, Skill Formation and Capital Accumulation : A Theoretical Analysis', *Keio Economic Studies, Forthcoming.*

Harris, J. R. and Todaro, M. P. (1970) : 'Migration, Unemployment and Development : A Two-sector Analysis', *American Economic Review, 60, 126-142.*

Jagarey, S. and S. Lahiri (2002) : 'Will Trade Sanctions Reduce Child Labour? The Role of Credit Markets', *Journal of Development Economics, 68(1), 137-156.*

Jayaraj, D. and Subramanian, S. (1997) : 'Child Labour in Tamil Nadu : A Preliminary Account of its Nature, Extent and Distribution', *Working Paper No. 151, Madras Institute of Development Studies, Chennai.*

Kulkarni, Manu (1983) : 'Match-making Children of Sivakasi', *Economic and Political Weekly, October 22.*

Majumdar, D. (1975) : 'The Urban Informal Sectors', *IBRD Staff Working Paper, 211, Washington.*

Ranjan, P. (1999) : 'An Economic Analysis of Child Labour', *Economics Letters, 64, 99-105.*

Ray, Ranjan (1998) : 'Analysis of Child Labour in Peru and Pakistan : A Comparative Study', *mimeo : Hobart, Australia.*

Romatet Emmanuel (1983) : 'Calcutta's Informal Sector : Theory and Reality', *EPW 18, December 10, 2115-2128.*

Swaminathan, Madhura (1997) : 'Do Child Workers Acquire Specialized Skills? A Case Study of Teenage Workers in Bhavnagar', *Indian Journal of Labour Economics, 40 (4), 829-39.*

Tokman, V. E. (1989) : 'Policies for a Heterogeneous Informal Sector in Latin America', *World Development, 17, (7). 1067-1076.*

Weeks, J. (1975) : 'Policies for Expanding Employment in the Informal Urban Sector of Developing Economies', *International Labour Reviews, 3(1), 1-13.*

Weiner, Myron (1996) : 'Child Labour in India : Putting Compulsory Primary Education on the Political Agenda', *Economic and Political Weekly, Vol. Nov. 9-16, pp.45-6.*

Weiner, Myron (1991) : 'The Child and the State in India : Child Labour and Education Policy in Comparative Perspective', *Princeton University Press, Princeton.*

Arthaniti 2 (1-2), 108-118 / 2003

When scrutiny reduces signaling incentives*

Debabrata Datta** and Jaideep Roy***

Abstract

In a non-discriminatory environment, a principal hires an agent for a specified period. The true skill of the agent is unknown to the principal but can be learned through informative signals. The precision of such signals depends positively on the level of scrutiny, undertaken by the principal and full scrutiny reveals the truth. The skill determines the future productivity and its knowledge is essential for the principal to decide upon whether to re-employ the agent or not, once the initial contract terminates. We show that full scrutiny is sub-optimal and hurts the principal. It is in the best interest of the principal to opt for significantly low levels of scrutiny during the probation even if the scrutiny is costless.

JEL Classification : C7, D8, J3.

Keywords : Scrutiny, signaling incentives, nondiscriminatory environment.

---

1. Introduction

In many work environments, the employer is not fully aware of the skill of a job applicant at the time of hiring. Nevertheless, if the employer believes that with some probability the applicant promises good returns in the future, she may take a chance by employing him for a stipulated period of time. Also, during this initial phase of employment (the duration of which may vary from one type of job to another), the employer enjoys the discretion to scrutinize the skill of the employee by on-job observation or by other forms of direct and indirect interaction. Such scrutiny becomes important for the employer as it enables her to gain knowledge regarding the true skill of the agent concerned in order to decide whether or not to reemploy him for the future. It is precisely this that induces employers to scrutinize their work force and increase efficiency. Nevertheless, in many instances scrutiny does not come without costs incurred by the employer. As a result the chosen degree of scrutiny depends crucially upon the benefits and the costs associated with it, notwithstanding the fact that if scrutiny were costless, one would expect employers to choose unbounded levels of it.

In this paper we show that even when the employer has costless scrutiny at her disposal and requires further knowledge to guarantee the skill of her workforce and

---

[*] We thank an anonymous referee for suggestions, which improved the clarity of the paper. Several comments from Pradeep Dubey are gratefully acknowledged. The paper also benefited from seminar presentation at the Department of Economics, University of Calcutta. The usual disclaimer applies.

[**] Department of Economics, Asutosh College, Calcutta University, 92 S. P. Mukherjee Road, Calcutta 700 026, India. Tel: +91 33 2465 3553, Fax: +91 33 2464 7596; Email: datta_d21@sify.com

[***] Department of Economics, Leon Kozminski Academy of Entrepreneurship and Management, Ul. Jagiellonska 59, 03 301 Warsaw, Poland. Tel: +48 22 701 8432, Fax: +48 22 811 2022; Email: jroy@wspiz.edu pl

thereby increase efficiency, she may choose to even reduce scrutiny to a negligible level. Indeed full scrutiny reveals the true skill of an employee, yet, there may be an efficiency loss due to lack of work incentives on part of the work force. In contrast, by choosing even a negligible degree of scrutiny she is able to learn the true skill with probability one and simultaneously achieve productive efficiency as under discriminatory environments.[1]

The environment studied is kept as simple as required. For an initial duration, a principal operating in a nondiscriminatory environment hires an agent whose skill is private information but can be learned publicly by use of scrutiny. Higher is the degree of scrutiny better is the information. The agent can be either high skilled or low skilled. If the agent is high skilled, the principal gains from reemploying him while if he is low skilled, reemployment hurts and therefore she does her best to avoid such unwanted events. The principal adheres to a level of scrutiny, which is common knowledge. Consequently, the agent chooses an unobservable effort level that determines current production. Finally, the initial hiring phase terminates and the principal decides upon her reemployment decisions.[2] We assume for simplicity that the agent works at zero current wages as his reservation payoff is set equal to zero. Motivation to work is driven solely by expected future returns from reemployment as in the literature on the economics of career concerns [see Holmstrom (1982) and Dewatripont et al. (1999)]. The driving force behind our result is as follows. Too much scrutiny makes the agent realize that the principal knows his true skill with almost certainty. In that case the agent knows that if he is high skilled he will be reemployed while if he is low skilled he will be evicted with very high probability and independent of his choice of current effort. Thus the agent exerts zero effort, no matter what is his true skill. On the other hand, with very low levels of scrutiny, a high skilled agent has an incentive to choose an effort level in order to signal his true skill to the principal. In equilibrium, one observes i) a very low level of scrutiny, which also includes its complete absence, ii) a low skilled agent once again shirks completely while a high skilled agent exerts positive effort in order to signal his true skill, and iii) at the time of reemployment the principal knows the true skill with probability one, which she would have, had she undertaken full scrutiny. Thus, too much scrutiny takes away the signaling incentives from the high skilled agent and so even if scrutiny is costless and helps the principal to make correct reemployment decisions, low levels of scrutiny becomes even more attractive for the principal as it increases current production without hurting her reemployment decisions.

---

[1] By choice of convention, we call an environment discriminatory if at the time of hiring, the pnncipal offers standard menus of different contracts for the same job in order to screen between agents with different skills. We restrict our principal from doing this, as in modern times discriminatory environments are rarely observed.

[2] Two comments are due with respect to the term "probationary" which highlight its importance to a firm. First, such probationary periods are recurring since at any point in time, a firm usually would have some employees under probation for future recruitment. Second, one can also think of such periods as a normal working penod where even a permanent employee is up for a possible promotion. Then efficiency during probation in our analysis is extended to general work efficiency, as such.

A similar motivation can be found in Dubey and Wu (2000). In their nondiscriminatory competitive environment, a principal appoints two agents to produce a specified quantity of output, say one thousand light bulbs. After production takes place, the principal chooses how many light bulbs would be checked in order to detect production errors. Finally the principal buys the produced one thousand bulbs from that agent for whom the number of detected errors from the chosen sample size is less. Such contracts are termed "competitive prizes" and the size of the chosen sample is the measure of scrutiny in their context. Of the two agents, one is high skilled while the other is low skilled with the classification that the probability of producing a defective bulb is lower for the high skilled agent. They show that even if scrutiny is costless, the principal may find it optimal to reduce the sample size significantly below the one thousand mark. This is because with full scrutiny, it becomes almost impossible for the low skilled agent to win the prize and knowing this, the high skilled agent also exerts minimum efforts to guarantee quality. On the other hand, with very low levels of scrutiny, each agent foresees a chance of winning. This in turn results in productive efficiency as an outcome of competition between the two agents, induced by a low level of scrutiny. We show that similar results can be obtained even without resorting to competitive contracts. Rather, with a single agent of unknown skill where the agent may wish to signal his true skill, scrutiny takes away signaling incentives from the high skilled agent and hurts current production.

The rest of the paper is structured as follows. In the next section we briefly lay out the simple model. Section 3 states and proves the result. The paper draws its conclusions in Section 4.

2. The model

A principal has hired an agent for a specified period of time to produce utility. We normalize payoffs and assume for simplicity that the wage rate offered to the agent during this initial probationary contracting period is zero. Let $\theta \in \{h, l\}$, be the true inherent skill or productive efficiency of the agent, which reflects his finite marginal cost of effort with $0 < h < l$. The true value of θ is not known to the principal and is held as equally likely by her (equally likely prior). Let $e \in R_+$ be the effort exerted by the agent with a cost function $c(e; \theta) = \theta e$. As usual, effort is not observable by the principal. If the agent of skill θ exerts an effort e, then the output produced during the probationary contracting period is $q = \varepsilon g(e)$ with $g(.)$ being a continuous and strictly increasing function and ε is a random variable whose expected value is 1. Since output is stochastic, it is not contractible. Upon termination of the probationary period, the principal faces the problem of taking the decision of whether to absorb the agent or lay him off. If the principal re-employs an agent with skill θ then she enjoys an expected future utility of $W(\theta)$ while the agent receives $V > 0$; where $W(h) > 0 > W(l)$ and $|W(l)| > |W(h)|$ We also make the assumption that $g(V/l) + W(l) \geq 0$ which translates to the fact that if a low skilled agent produces efficiently and the principal re-employs him in the future, her payoff over the entire horizon is non-negative. Eviction of the agent yields zero future utility to all parties involved. Clearly if the expected future utility of the principal from reemploying the agent is not negative the agent is reemployed. Otherwise, the principal refuses employment.[3] We assume that all players are risk - neutral.

During the probationary period, the principal has the option of supervising the agent to detect the true skill in order to make the optimal re-employment decision. Let $\sigma \in [0,1]$ denote the level of scrutiny where $\sigma = 0$ signifies zero scrutiny while $\sigma = 1$ implies full scrutiny. Let s be an informative signal regarding the true value of θ which can be observed by the principal[4] at each level of scrutiny defined by the following information matrix.

$$\begin{array}{lll} & \theta = h & \theta = l \\ s = 1 & p(\sigma) & 1 - p(\sigma) \\ s = 0 & 1 - p(\sigma) & p(\sigma) \end{array}$$

where p is the probability of receiving a signal $s = 1$ condition on the true skill being $\theta = h$ and the probability of receiving the signal $s = 0$ condition on the true skill being $\theta = l$. Thus $p(.)$ measures the degree of precision of the signals and is a function of the level of scrutiny σ with $p(0) = 1/2$; $p(.)$ being continuous and strictly increasing with $p(1) = 1$. In other words complete absence of scrutiny yields no additional information and as the degree of scrutiny is increased the principal moves closer to the truth. We shall be interested in finding the optimal level of scrutiny chosen by the principal. This analysis is done extensively in the following section.

---

[3] To be sure, observe that while the payoff g to the principal during the probationary period depends directly on effort, future utility W depends solely on skill. This is a modeling simplification where W must be thought of as a reduced form value function. In principle, future utility to the principal certainly depends on effort exerted by the re-employed agent and consequent output, stochastic in nature (as also on any other factor that affects output). However, since after the probationary period is over, the principal would simply ask for an output that provides exactly the reservation payoff to the agent, we have, $W(h) > W(l)$.

[4] Consider the following example. An employer decides how many times she will visit the work place and converse with the agent. This is her choice of σ. Each time the employer has a conversation with the agent, the agent may impress her or disappoint her (independent of the agent's inherent skills). If the number of times she gets impressed exceeds the number of disappointments, then we would call this a signal $s = 1$ received by the employer, while if the opposite happens we say that $s = 0$. The question that arises now is how reliable are these signals? Clearly, by invoking large sample arguments, higher values of σ increase this reliability. It is exactly this notion that is reflected in the information matrix.

3. Analysis

For every $\sigma \in [0,1]$ chosen initially by the principal we begin by defining the game $G\sigma\,(e, x)$ induced by the environment and played in three stages between the principal and her agent.

Stage 1. Nature chooses the true skill θ and the principal receives a public signal $s \in \{0,1\}$ whose precision is given by $p(\sigma)$.

Stage 2. The agent chooses $e \in R_+$ which is observed by the principal.

Stage 3. The principal chooses $x \in \{0,1\}$ where $x = 0$ implies that the agent is evicted while $x = 1$ implies that he is reemployed.

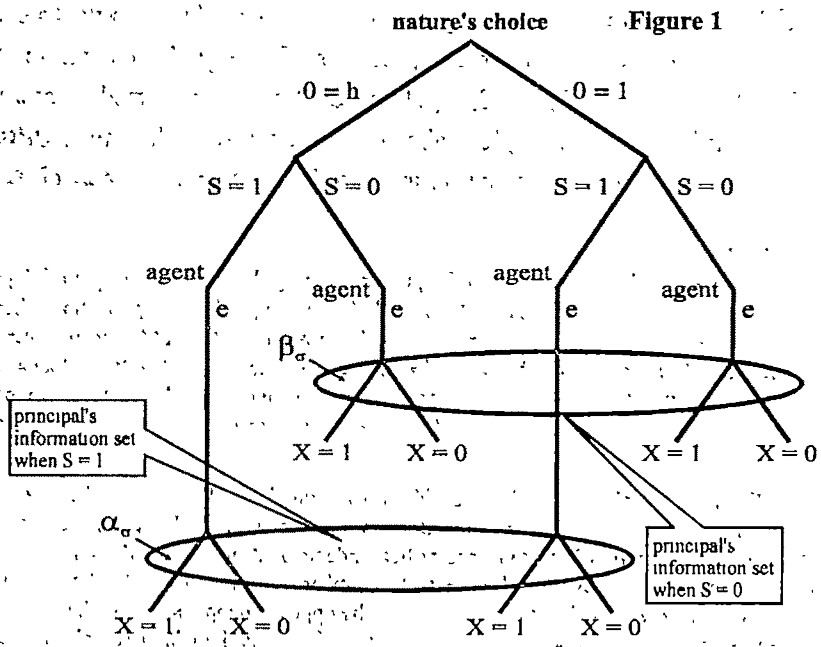
The payoffs of the game $G\sigma\,(e, x)$ are as follows. Suppose nature's choice of the true skill is θ; the agent exerts an effort e and the principal chooses the indicator x. Then the payoff to the principal is

$$u_\sigma\,(e, x|\,\theta\,) = g(e) + xW(\theta\,) \tag{1}$$

while the payoff to the agent is

$$V_\sigma(e,x|\theta) = -\theta e + xV \tag{2}$$

Figure 1: The game $G\sigma\,(e, x)$



As before, we assume that the reservation payoff of the agent is zero and that everything above is common knowledge. Such a game $G\sigma\,(e, x)$ is depicted in figure 1.

In what follows, for every given value of $\sigma \in [0,1]$, we find the Perfect Bayesian Equilibria of Gσ (e, x) and then find the optimal value σ^* chosen by the principal and consequently the optimal game Gσ^* (e, x).

The following theorem proves that it is not only sub optimal for the principal to undertake full scrutiny but also that complete absence of scrutiny can be observed in equilibrium, that is,

Theorem 1 : $\sigma^* \in [0, \sigma_l)$ *for some* $\sigma_l < 1$. *Consequently, if scrutiny involves costs, no matter how small, it is in the best interest of the principal to set* $\sigma^* = 0$.

Proof. Consider figure 1 which depicts the game Gσ (e, x) when α_σ is the posterior belief of the principal that $\theta = h$ when she receives the signal $s = 1$ and observes the effort exerted by the agent, while β_σ is the posterior belief of the principal that $\theta = h$ when she receives the signal $s = 0$ and observes the effort exerted by the agent. Note that before observing the exerted effort, if the principal chooses some $\sigma \in [0,1]$, her updated priors are

$$Pr[\theta = h|s = 1] = Pr[\theta = l|s = 0] = p(\sigma) \text{ and}$$
$$Pr[\theta = l|s = 1] = Pr[\theta = h|s = 0] = 1 - p(\sigma)$$

Since under no circumstances can e exceed V/l whenever $\theta = l$, for all $\sigma \in [0,1]$ we have

$$\alpha_\sigma(e) = p(\sigma) \text{ if } e < V/l$$
$$1 \text{ otherwise.}$$

and

$$\beta_\sigma(e) = 1 - p(\sigma) \text{ if } e < V/l$$
$$1 \text{ otherwise.}$$

Notice that any effort signal is either fully informative or has no information at all as is evident from the $\alpha_\sigma(.)$ and the $\beta_\sigma(.)$ functions. It is also clear from these functions that we make the usual tie - breaking assumption that if by choosing a positive effort an agent receives a payoff equal to zero, then the agent decides to shirk completely. This implies that the effort chosen by a low skilled agent always falls short of V/l.

Suppose $\sigma = 1$. If $\theta = h$, then the agent knows that no matter what is e, the principal knows for sure that $\theta = h$ and therefore her expected payoff from re-employing the agent is $W(h) > 0$. Since then the optimal strategy of the principal is $x = 1$ that guarantees the agent a payoff of $V > 0$, the optimal choice of effort from the high skilled agent is $e = 0$. If on the other hand $\theta = l$, the agent knows that no matter what is e, the principal knows for sure that $\theta = l$ and therefore her expected payoff from re-employing the agent is $W(l) < 0$. Since then the optimal strategy of the principal is $x = 0$, the optimal choice of effort from the low skilled agent is also $e = 0$. Thus, in equilibrium the game G_1 (e, x) generates an expected payoff of $\frac{W(h)}{2}$ to the principal and the agent exerts zero effort during the initial contracting period no matter what is his true skill.

Now suppose the principal has chosen some $\sigma \in [0,1)$

Case 1. $s = 1$

It is easy to see that whenever

$\alpha_\sigma(e)\, W(h) + \{1 - \alpha_\sigma(e)\} W(l) \geq 0,$

the agent knows that $x = 1$. Therefore, if σ is such that

$p(\sigma) W(h) + \{1 - p(\sigma)\} W(l) \geq 0,$

the best response of the agent is to choose $e = 0$ for any $\theta \in [h, l]$ as implied by the α_σ (.) function. Rewriting the above inequality, we get

$$p(\sigma) \geq \frac{-W(l)}{W(h) - W(l)}$$

Now take any $\sigma \in [0, \sigma_1)$ where σ_1 is such that $p(\sigma_1) = \frac{-W(l)}{W(h) - W(l)}$. In this case $p(\sigma) W(h) + \{1 - p(\sigma)\} W(l) < 0,$

and therefore the high skilled agent would like to give a signal to the principal through his choice of effort e such that $\alpha_\sigma(e) \geq p(\sigma_1)$. Since the low skilled agent will never exert an effort greater or equal to V/l, the higher skilled agent would optimize by choosing $e = V/l$ Anticipating this, the low skill agent would put zero effort.[5] The expected payoff to the principal from choosing a $\sigma \in [0, \sigma_1)$ is therefore

$$\frac{g(v/l) + W(h)}{2}$$

Case 2. $s = 0$

Similarly, whenever

$\beta_\sigma(e)\, W(h) + \{1 - \beta_\sigma(e)\}\, W(l) \geq 0,$

the agent knows that $x = 1$. Therefore if σ is such that

$[1 - p(\sigma)] W(h) + p(\sigma) W(l) \geq 0,$

---

[5] The argument is like in a model of Bertrand pricing with differentiated marginal costs of production

the best response of the agent is to choose $e = 0$ for any $\theta \in \{h, l\}$ as implied by β_σ (.) function. Rewriting the above inequality, we get

$$p(\sigma) \leq \frac{W(h)}{W(h) - W(l)}$$

Now take any $\sigma \in [\sigma_0, 1]$ where σ_0 is such that $p(\sigma_0) = \frac{W(h)}{W(h) - W(l)}$. Then invoking the argument as in case 1 it can be shown that the high skilled agent chooses $e = V/l$ and the low skilled agent chooses $e = 0$. Once again the expected payoff to the principal is $\frac{[g(v/l) + W(h)]}{2}$. Notice that since $|W(l)| > |W(h)|$ we have $\frac{-W(l)}{W(h) - W(l)} > \frac{1}{2}$ and therefore $\sigma_1 > 0$. Also by the same assumption we have $\frac{W(h)}{W(h) - W(l)} < \frac{1}{2}$. Thus, by continuity of p (.) and the fact that $p(0) = \frac{1}{2}$, it follows that $\sigma_0 = 0$. This guarantees the existence of $\sigma \in [0, \sigma_1)$ such that in the game $G_\sigma(e, x)$, the principal receives a payoff equal to $\frac{[g(V/l) + W(h)]}{2}$ in expected terms, independent of the signal s. Clearly this is the highest expected payoff obtainable by the principal by choice of σ. To see this it is sufficient to make the following observations:

i) For any $\sigma \in [0, \sigma_1)$, the high skilled agent exerts a positive effort equal to V/l no matter what is the realization of the signal s and at the end of the initial contracting period, the principal learns the true skill with probability 1 and ii) the low skilled agent always exerts zero effort. Thus the choice set $[0, \sigma_1)$ clearly dominates the choice set $[\sigma_1, 1]$. To see this, take any $\sigma \in [\sigma_1, 1]$.

The expected payoff to the principal from choosing such a σ in equilibrium is $\frac{1}{2}[p(\sigma)W(h) + (1 - p(\sigma)W(l)] + \frac{1}{2}\left[\frac{g(V/l) + W(h)}{2}\right]$ and therefore what remains to be shown is that for such a $\sigma \in [\sigma_1, 1]$ we have

$$p(\sigma)W(h) + \{1 - p(\sigma)\}W(l) < \left[\frac{g(v/l) + W(h)}{2}\right]$$

Suppose not. That is suppose

$$p(\sigma)W(h)+\{l-p(\sigma)\}W(l)\geq\left[\frac{g(v/l)+W(h)}{2}\right]$$

Rewriting the above inequality, we get

$$p(\sigma)\geq\frac{g(V/l)+W(h)}{2[W(h)-W(l)]}+\frac{-W(l)}{W(h)-W(l)}$$

Since $\frac{-W(l)}{W(h)-W(l)}>\frac{1}{2}$

and $g(V/l)+W(l)\geq 0$ by assumption, from this it follows that

$$\frac{g(V/l)+W(h)}{2[W(h)-W(l)]}\geq\frac{1}{2}$$

This in turn implies that for our supposition to be true, we need $p(\sigma) > 1$ which is impossible. Also from the above it is easy to see that $\sigma^* \neq 1$. The rest of the theorem follows trivially and requires no proof. ■

The idea behind obtaining the above result is as follows and deserves some consequent remarks. Firstly, full scrutiny spoils the principal's show. This is because full scrutiny reveals the truth and thereby the principal reaches her reemployment decision, independent of what is the observed effort. The high skilled agent realizes that his skill has become common knowledge and therefore foresees guaranteed future contracting. He therefore exerts no costly effort. The low skilled agent on the other hand faces certain eviction and therefore also chooses to shirk. So with full scrutiny the principal expects to make her reemployment decision without any uncertainty but fails to generate any positive effort from both types of agents during the initial contracting period. If the principal reduces the scrutiny from $\sigma = 1$ the high skilled agent may fear that he shall be evicted by mistake. This is particularly so when the prior of the principal that the agent is high skilled is sufficiently low (which is whenever $s = 1$ and the level of scrutiny is sufficiently low or whenever $s = 0$ and scrutiny is at any level, barring the situation when $\sigma = 1$).

Thus in all such cases, the high skilled agent is strictly better off by exerting the minimum effort of V/l (and earn a positive surplus) as then the principal will be certain that the agent is of high skilled in which case she reemploys the agent. This ability of signaling one's true skill by choice of effort was precisely absent under full scrutiny. In all such cases the low skilled agent finds it optimal to shirk as no effort, which ensures him reemployment, also earns a positive payoff. By continuity of the problem, the above intuition holds for all levels of scrutiny close to 1.

We conclude this section with the following remark. Even with an arbitrarily small cost of scrutiny, the optimal level of scrutiny becomes equal to zero. Moreover even if scrutiny is costless, full scrutiny is never desirable. It is also the case that as $|W(l)|$ approaches $|W(h)|$, the value of σ_1 as defined above tends to zero. This shrinks the region for optimal scrutiny and only significantly small levels of scrutiny is chosen in equilibrium. Finally, the fact that $\sigma^* < 1$ is independent of the assumption that $g(V/l) + W(l) \geq 0$. This assumption is used to ensure that the optimal set of scrutiny shrinks as $|W(l)|$ approaches $|W(h)|$.

4. Concluding remarks

Scrutiny is a widespread practice to provide the buyer with the means to gain knowledge regarding the quality of goods and services. For this reason, we presume that our result may have implications in various economic institutions, some of which we discuss below.

1. We have extensively studied our fundamental idea in case of the labor market where the employer to ensure productive efficiency typically uses scrutiny. In this sense, the decision of optimal scrutiny becomes fundamental in the efficient operation of labour market. With the help of a simple model where scrutiny requires no additional costs, we show that although very high levels of scrutiny enables the employer to learn about the quality of her work force, too much of it may also hamper effort incentives. Rather, scrutiny to a bare minimum serves the twin purposes of revealing information as well as improving production, whenever there is chance of some future prize for the agent.

2. This basic finding may also explain why often buyers in the residential market areas (for example retail household markets, where most of the buyers are regular customers) leave the task of choosing the fruits or vegetables to the sellers, although with little effort they can do this job themselves, while in market where stray buyers outnumber regular buyers, the buyers are observed to check the products fastidiously. The possibility of future purchase in the former market generates the signaling incentive for the seller, which the buyers realize that self- scrutiny will take away. On the contrary, in stray buyer markets, signaling has relatively low value.

3. Another interesting implication of our theoretical framework may be in agriculture. It has long been argued by many economists and policy- makers that the institution of absentee landlordism is a social evil and should therefore be eliminated. The argument has been that by staying away from his land, the landlord is unable to monitor various activities carried on by employed laborers and farm tenants. As a result, such absence from one's own land hampers agricultural output. Nevertheless, as the institution is highly widespread, the popular explanation has been that landlords stay away from their land simply because the opportunity costs (in terms of other

sources of income, leisure and quality of life) of residing close to the farm is excessively high for them. From what we show, it may be the case that some landlords strategically wish to stay away and commit to a reduced level of scrutiny on his workers, thereby increasing production without hurting his reemployment decisions and future profits from agriculture.

4. The finding may also have important policy implication for public administration. The bureaucracy usually has an in-built system of high level of scrutiny. Our paper shows that this high level of scrutiny is Pareto- sub optimal (doubly so if scrutiny is costly) and can be avoided, provided of course that future benefits, accruable to the suppliers of goods and services to the government are common knowledge and the administration can exercise the authority of denying that future benefit to the unproductive workers or supplier of service.

References

Dewatripont, Mathias, Ian Jewitt and Jean Tirole, (1999): 'The Economics of Career Concerns, Comparing Information Structures', *Review of Economic Studies, 66, 183-98*

Dubey, Pradeep and Chen Wei Wu, (2000): 'Competitive Prizes', *Ph.D. Dissertation, Dept. of Economics, SUNY Stony Brook, USA*

Dubey, Pradeep and Chen Wei Wu, (2001): 'Competitive Prizes: When Less Scrutiny Induces More Effort', *Journal of Mathematical Economics, 36(4), December 2001: 311-36.*

Holmstrom Bengt, (1982): 'Managerial Incentive Problems: A Dynamic Perspective', *in Essays in Economics and Management in Honor of Lars Wahlbect, Swedish School of Economics.*

Book Review

Sangeeta Kamat
Development Hegemony: NGOs and the State in India
Oxford, 2002

One of the central moments of contemporary critical-intellectual work—broadly within the field of what has come to be designated as Postcolonial theory (distinguished from other streams of post-colonial studies in terms of its epistemological moorings in post-structuralist, anti-essentialist modes of analysis)—has been its inauguration of the *discourse-analysis* of 'Development'. Understandably so, since postcolonial experience has been uniquely articulated, shaped and *stabilized* within the dominant, overwhelming thematic of development.

By the late 70s, the incidence of gross inequality, persisting poverty and debilitating 'dislocations' in the cultural and social sphere in the Third World shattered the illusion or myth of development. Resistance to development projects multiplied. At the same time, the 'withering away' of the 'developmentalist state in the resurgent 'neo-liberal' regime released a space for public intervention that came to be characterized by two different and contradictory tendencies. On the one hand, the failure of macro-projects of development in transmitting its benefits to the population at the lower strata led to 'targeting' of groups and implementation of community development projects with emphasis on grassroots initiatives. The NGOs and other voluntary organizations stepped in as the main agencies of this project. On the other hand, the erosion of Left (party) politics (as the resistant space of politics) led to the dispersion of the political force, particularly the youth, who then sought to reorganize and reengage themselves in grassroots work in the form of small but numerous "non-party political formations", committed to work among the urban and the rural poor. While the former act as the functionary of the discourse of (decentralized/participatory) development, the latter is characterized by their contestation of the discursive practices of development and by their efforts at politicization of the space of development. Slowly, parallel to the discourse of participatory development, another discourse of "new social movements" (meaning the space of political grassroots mobilization) came to take shape. Sangeeta Kamat's book takes up this issue at its location in the interface of these two discourses.

Kamat belongs to that strand of critical thinking that has attempted to reveal the project of development as a powerful regime of Foucaultian *power/knowledge*. This particular school of thought, known as post-development, has held the view that the 'discourse' of development has always represented as a technical issue what was essentially a political one, thereby erasing the power relations (and hence, its associated violence) that punctuate the space of development. The tribulations of the development experience in the Third World have not only shattered the 'myth', but also inaugurated a space of resistance against development practices. The post-developmentalists have generally tended to locate the locus of resistant politics in the 'new social movements' (Narmada Bachao Andolan in India, the Zapatista movement in Mexico etc.) characterized by its 'grassroots' nature and its articulation of local, culture-specific and 'authentic' aspirations. However, it is the rootedness of these struggles in the 'local' that promise to usher in a new and *universal* language of resistance. It is by being 'local' that these movements seek to escape any universalizing project of the elite. Also, these movements harbor a political space largely autonomous of the space of state-politics. The third world state, in post-development theory, has been unequivocally treated as the main functionary of the discourse of development. The relative autonomy of these movements from the state, therefore, carries with it the potentiality of engendering a space of politics that can resist and contest the discursive practices of development.

Kamat's point of departure is the *possibility* or rather the *sustainability* of this space of resistance. ,What distances her from other post-development theorists and renders her work a unique contribution to this growing body of work is that she seeks to problematize resistance itself. Through detailed analysis of a specific grassroots movement, she reveals the nuances of the space of resistance. Though she believes that the grassroots organizations represent one of the "primary sites of resistance", she contests the assumed autonomy of these spaces from the state (or the 'idea' of the state, as she puts it) and hence from the discursive practices of development. Kamat, however, does not focus on the "new social movements", but rather on the "non-party political formations" which, she claims, despite their avowed solidarity with the "new social movements", exhibit a strong tendency of getting incorporated in the 'community development projects'—a process which Kamat refers to as the "NGO-ization of grassroots movements".

Kamat illustrates this process though the first-hand account of a particular grassroots mobilization in a Maharashtrian tribal village. This particular grassroots initiative, in response to the economic compulsions of landless tribal peasants on the one hand and its awareness of the futility and debilitating effects of development projects on the other hand, developed a dual-structure organization, one organ of which (the *Seva Sansad*) took up and implemented 'income-generation schemes' of the government for immediate economic reprieve of the poor, while the other (the *Shramik Hakk Sanghatna*) acted as the political foil to the anti-political discursive practices of development and sought to mobilize the peasants around political issues that emphasized the state structure and unequal social relations of production.

What Kamat observes is the reproduction of the state ideology not only in the discourse of the *Sansad* (that is to be expected), but also in the *Sanghatna*'s discourse of anti-state, anti-development politics. Deploying the Gramscian category of 'hegemony', Kamat characterized the situation as one of state ideology attaining the 'status of hegemony'—in the sense of extending its sway over the civil society. More precisely, she identifies those central moments that characterize both the discourses of the state and the space of resistant politics—(a) separation of the economic from the political, (b) commodity fetishism, (c) reification of the state and (d) individualization of social relations of production. She illustrates these moments through her account of one major development project undertaken by the *Sansad/ Sanghatna*—the dairy project. Kamat, in fact, presents these moments as characterizing the "NGO-space", in opposition to which she posits the "authentic" resistant political space.

Kamat maintains that organizational structures like the source of funding (external), nature of the staff (professional) and state patronage by themselves do not tell us about the nature of politics of a grassroots organization. It's through the analysis of "language, practices and everyday political choices" faced by these organizations that one can uncover the real orientation of these organizations vis-a-vis the elite projects of "development, capitalism, and neoliberalism".

One of the central problems with Kamat's book is its epistemological vagueness. Though she claims to employ a post-structuralist analytical framework, in reality, she eclectically combines the elements of the methodologies of Gramsci, Foucault and Marx, without any attempt at synthesis. Foucault, as it is, does not marry well with Marx. Gramci's categories are often in contradiction with orthodox Marxist framework. And Foucault and Gramsci cannot be synthesized in an essentialist terrain. All these are well recognized at the theoretical level. But she offers no clear epistemological position on these issues in her work.

Despite these weaknesses, however, her work is a novel intervention in the discourse of "grassroots movements" which, one might say, she has impregnated with much awaited informed critical insights.

– Rajesh Bhattacharya

INSTRUCTIONS TO AUTHORS:

- The Arthaniti considers manuscripts for publication on the understanding that the submitted paper represents an original work not previously published, or being considered for publication elsewhere. The Journal has the copyright to all its published articles.
- Manuscripts accepted for review are blind refereed by two or more scholars. In this context, the authors are advised to write papers in a way to keep their identity from referees. The first page of the text should not reveal the name(s) of the author(s).
- Three hard copies of the manuscript along with an abstract in electronic form (not exceeding 100 words) should be sent to Dr. Kanailal Basak, Editorial Secretary, Arthaniti, Dept. of Economics, University of Calcutta, 56A, B. T. Road, Kolkata 700 050, India.
- Although there is no submission fee or page charge, each manuscript should be accompanied by a processing fee, which is Rs. 100 for authors within India and US $ 10 for authors outside India. Demand drafts must be made payable to "ARTHANITI".
- Manuscripts should be in English and between 4,000 and 8,000 words (15-30 pages). They should be double- spaced and typed on standard letter-size (11 x 8 5) paper.
- The cover page should include the author's name, institutional affiliation, and the title of the paper Please include the contact person's full address, phone, fax numbers and e-mail address. Include the name(s) of any sponsor(s) of the research contained in the paper, along with grant number(s) Supply an abstract of up to approximately 100 words for all articles. It should also contain at least one classification code according to the JEL Classification and five keywords.
- Only essential mathematical notations should be used All equations and statistical formulae should be neatly typed and numbered consecutively throughout the text in Arabic numerals
- Footnotes should be kept to a minimum and numbered consecutively throughout the text with superscript Arabic numerals
- References should be quoted in the text as name and year within brackets and listed at the end of the paper alphabetically. Where reference is made to more than one work by the same author published in the same year, identify each citation in the text as follows· (Johansen, 1988a), (Johansen, 1988b). Where references include three or more authors the form (Johansen et al., 1988) should be used in the text. All references must be complete and accurate. Online citations should include date of access If necessary, cite unpublished or personal work in the text but do not include it in the reference list. References should be listed in the following style. Johansen, S. (1988): Statistical analysis of cointegrating vectors, *Journal of Economic Dynamics and Control* 12, 231-254. World Bank. (1994) *Adjustment in Africa· Reforms, Results and the Road Ahead.* Oxford University Press, New York.
- Tables should be numbered consecutively in the text in Arabic numerals and printed on separate pages
- Illustrations. Supply each illustration on a separate sheet, with the lead author's name and the figure number, with the top of the figure indicated, on the reverse. Supply original photographs, photocopies or previously printed material will not be used Line artwork must be high-quality laser output (not photocopies). Tints are not acceptable; lettering must be of a reasonable size that would still be clearly legible upon reduction, and consistent within each figure and set of figures. Supply artwork at the intended size for printing.
- Final version of the papers must be submitted on floppy diskette in MS Word 2000 / Word Perfect 6.0 along with two hard copies
- Final proofs will also usually be sent to the author for checking. This stage is to be used only to correct errors that may have been introduced during the production process. Prompt return of the corrected proofs. preferably within a week of receipt, will minimize the risk of the paper being held over to a later issue.
- A copy of the journal will be provided, free of charge, to each author.